C.D. Pagani S. Salsa

ANALISI MATEMATICA

Volume 1

Milano • Parigi • Barcellona
1995

INDICE

Elementi di teoria degli insiemi 1

1. NOZIONI DI LOGICA MATEMATICA

Presentiamo in questo paragrafo alcuni elementi di logica matematica. Scopo di questa disciplina è analizzare e formalizzare i metodi corretti di ragionamento. Le ricerche in questo campo hanno avuto origine all'inizio di questo secolo, stimolate fra l'altro dalla scoperta delle antinomie o paradossi (contraddizioni logiche) nella teoria cantoriana degli insiemi. Si è allora sviluppata un'opera vasta di revisione dei fondamenti della matematica, che ancora non si può dire conclusa. Al di là delle questioni teoriche che l'hanno originata, quest'opera ha già prodotto una ricaduta feconda di applicazioni concrete, particolarmente nello studio dei linguaggi dei calcolatori.

Noi ci limiteremo a discutere quel poco di logica che ci servirà per lo sviluppo dell'Analisi.

Supporremo, nel presente capitolo, una certa familiarità dello studente con i numeri naturali, relativi, razionali e reali. Ritorneremo più avanti su questi argomenti in maniera organica.

1.1 Logica delle proposizioni. Connettivi logici

Nella lingua parlata gli elementi costitutivi sono le *proposizioni* (o affermazioni o enunciati); lo stesso avviene in matematica. In questo paragrafo porremo gli enunciati tra virgolette, che potremo eliminare in seguito qualora non sorgano ambiguità. Esempi di enunciati sono:

"Socrate è un uomo";
"Pitagora mangia la mela";
"$7 \cdot 3 = 21$";
"3 è un numero naturale";
"$2 + 2 \geq 5$".

Caratteristica degli enunciati è che possiamo attribuire loro un valore di verità: *vero* (V) o *falso* (F). Ad esempio il terzo dei precedenti enunciati è vero mentre il quinto è falso.

Indicheremo gli enunciati con lettere come p, q, r ecc.

Data una proposizione p si può costruire la sua *negazione* che indichiamo con "$\sim p$" (si legga "non p"). Il simbolo $\sim$ sta ad indicare quindi l'avverbio *non*. Ovviamente, se p è vera, $\sim p$ è falsa e viceversa. Ad esempio, se p è l'enunciato (falso):

"3 è un numero naturale pari",

allora "$\sim p$" è l'enunciato (vero):

"3 non è un numero naturale pari".

Le proposizioni possono essere legate tra loro dando luogo a proposizioni più complesse. Nel linguaggio comune il collegamento viene effettuato per mezzo delle congiunzioni. In logica matematica i termini di collegamento si chiamano *connettivi logici*. Anche il simbolo $\sim$ (non) è considerato un connettivo.

Gli altri sono i seguenti:

la *congiunzione*, che corrisponde alla congiunzione "*e*". Essa viene indicata col simbolo $\wedge$ ma anche semplicemente con la lettera *e*.

La *disgiunzione*, che corrisponde alla congiunzione "*o*", nel senso del "vel" latino (contrapposto ad "aut"). Essa viene indicata col simbolo $\vee$ ma anche semplicemente con la lettera *o*.

La *implicazione*, che corrisponde alla locuzione "se ... allora"; viene indicata col simbolo $\Longrightarrow$.

La *doppia implicazione*, che corrisponde alla locuzione "se e solo se"; viene indicata col simbolo $\Longleftrightarrow$.

Usando i connettivi logici e le parentesi possiamo formare enunciati composti come i seguenti:

$$p \Longrightarrow (q \Longrightarrow r) \ , \ (p \wedge q) \Longleftrightarrow r \ , \ (\sim p) \Longrightarrow (q \wedge r) \quad \text{ecc.}$$

Per evitare di sovraccaricare gli enunciati composti con troppe parentesi si conviene di introdurre tra i connettivi un ordine gerarchico esattamente come si fa in aritmetica per le quattro operazioni $(+, -, \cdot, :)$.

Ad esempio la scrittura $2 \cdot 3 + 5$ significa $(2 \cdot 3) + 5$ e non $2 \cdot (3 + 5)$. Si suol dire che moltiplicazione e divisione "legano" più di addizione e sottrazione.

Analogamente, introduciamo tra i connettivi il seguente ordine dove ciascun connettivo "lega" più dei successivi:

$$\sim, \wedge, \vee, \Longrightarrow, \Longleftrightarrow \ .$$

Con questa convenzione invece di scrivere

$$((\sim p) \Longrightarrow (\sim q)) \Longleftrightarrow (p \vee (\sim r))$$

potremo scrivere semplicemente

$$\sim p \Longrightarrow \sim q \Longleftrightarrow p \vee \sim r$$

1.2 Tavole di verità

Quando due proposizioni vengono legate insieme usando i connettivi logici, il valore di verità della proposizione che ne risulta dipende dal valore di verità delle proposizioni componenti, secondo le seguenti regole o *tavole di verità*:

Per il simbolo $\sim$:

p	$\sim p$
V	F
F	V

Per gli altri:

p	q	$p \wedge q$	$p \vee q$	$p \Longrightarrow q$	$p \Longleftrightarrow q$
V	V	V	V	V	V
F	V	F	V	V	F
V	F	F	V	F	F
F	F	F	F	V	V

A parole:

"$p \wedge q$" è vera se p e q sono entrambe vere; falsa negli altri casi
"$p \vee q$" è falsa se p e q sono entrambe false; vera negli altri casi
"$p \Longrightarrow q$" è falsa se p è vera e q è falsa; vera negli altri casi
"$p \Longleftrightarrow q$" è vera se p e q sono entrambe vere o entrambe false; falsa negli altri casi.

Ad esempio la proposizione

"$(2 + 2 \geq 5) \wedge$ (3 è un numero naturale)"

è falsa perché la prima proposizione è falsa. Analogamente:

"$(2 + 2 \geq 5) \vee$ (3 è un numero naturale)" è vera
"$(2 + 2 \geq 5) \Longrightarrow$ (3 è un numero naturale)" è vera (!!)
"$(2 + 2 \geq 5) \Longleftrightarrow$ (3 è un numero naturale)" è falsa.

Lo studente avrà notato che l'uso del connettivo $\Rightarrow$ non corrisponde propriamente alla locuzione "se ... allora" del linguaggio comune. Infatti, quando noi usiamo tale locuzione, riteniamo che le due proposizioni componenti p e q abbiano una correlazione causale o qualche altro tipo di legame. Invece in matematica la verità della proposizione "$p \Longrightarrow q$" dipende solo dai valori di verità di p e q, anche se tali proposizioni non sono correlate fra loro.

L'origine della tavola di verità di "$p \Longrightarrow q$" può essere spiegata intuitivamente come segue.

Esaminiamo la frase

"se c'è il sole faccio una passeggiata".

Ciò è equivalente all'alternativa

"non c'è il sole oppure faccio una passeggiata".

Si noti che il fatto che non ci sia il sole non implica alcuna decisione riguardo alla passeggiata.

Dunque, nel linguaggio comune, dire:

"$p \Longrightarrow q$"

è equivalente a dire

"$\sim p \vee q$" .

La tavola di verità di "$\sim p \vee q$" è la seguente:

p	q	$\sim p$	$\sim p \vee q$
V	V	F	V
V	F	F	F
F	V	V	V
F	F	V	V

Tale tavola coincide con quella di "$p \Longrightarrow q$".

Dunque, se riteniamo "corretta" la tavola di verità per il connettivo $\vee$, dobbiamo ritenere corretta anche quella indicata per l'implicazione.

Osserviamo infine che l'implicazione "$p \Longrightarrow q$" spesso si legge "p è condizione sufficiente perché valga q" oppure "q è condizione necessaria perché valga p". La doppia implicazione "$p \Longleftrightarrow q$", che esprime l'equivalenza logica delle due proposizioni p e q, si legge: "condizione necessaria e sufficiente perché valga p è che valga q".

1.3 Tautologie e regole di deduzione

Nella dimostrazione di un teorema o in ogni tipo di argomentazione logica si fa uso continuo di *regole di deduzione* (o di inferenza), cioè di regole logiche che precisano il corretto modo di ragionare (*).

(*) Preferiamo, a questo livello, ritenere intuitivo il concetto di dimostrazione.

Esaminiamo un classico esempio di argomentazione valida.

1) Socrate è un uomo.
2) Se Socrate è un uomo allora Socrate è mortale.
3) (Allora) Socrate è mortale.

La sequenza 1), 2), 3) è una deduzione corretta dell'enunciato 3) dagli enunciati 1) e 2); cioè è una dimostrazione che Socrate è mortale.

Osserviamo subito che 3) *non* si può dedurre solo da 2). In termini più generali, da "$p \Longrightarrow q$" *non* si può dedurre q.

Per rendersi meglio conto di questo fatto consideriamo l'enunciato

"se 3 è un intero pari allora 4 è un intero dispari".

Anche se tale enunciato è vero non possiamo certo dedurre che 4 è dispari.

Osserviamo inoltre che la validità della deduzione di 3) da 1) e 2) *non* dipende dal valore di verità di 1) e 2). Infatti, anche la seguente deduzione è corretta:

1) Socrate è un uomo.
2') Se Socrate è un uomo allora è immortale.
3') (Allora) Socrate è immortale.

Su che cosa è fondata la correttezza delle deduzioni precedenti?

Pensate di far eseguire la deduzione 1), 2), 3) da un elaboratore nella cui memoria siano inserite (come ipotesi) gli enunciati 1) e 2). Esso non sarà mai in grado di *dedurre* l'enunciato 3) a meno che non lo forniate della regola seguente: "se in memoria si trovano gli enunciati p e "$p \Longrightarrow q$" allora si può inserire anche q".

Cerchiamo di capire l'origine di questa regola di deduzione.

Consideriamo l'enunciato seguente

(MP) "$p \wedge (p \Longrightarrow q) \Longrightarrow q$"

e la sua tavola di verità:

p	q	$p \Longrightarrow q$	$p \wedge (p \Longrightarrow q)$	MP
V	V	V	V	V
F	V	V	F	V
V	F	F	F	V
F	F	V	F	V

Qualunque siano i valori di verità di p e q, il valore di verità di (MP) è sempre V.

Essendo (MP) sempre vera, è naturale considerare q come conseguenza logica di p e "$p \Longrightarrow q$" ovvero formulare la seguente regola di inferenza che si chiama *modus ponens*:

da p e "$p \Longrightarrow q$" si deduce q .

Dal punto di vista pratico il modus ponens opera così: *se p è vera e si vuole mostrare che q è vera, si procede mostrando che "$p \Longrightarrow q$" è vera.*

Gli enunciati come (MP), veri qualunque sia il valore di verità delle proposizioni componenti, prendono il nome di **tautologie**.

Un'altra importante tautologia è la seguente, nota come *principio di contrapposizione*:

$$\text{(PC)} \qquad \text{“}p \Longrightarrow q \iff \sim q \Longrightarrow \sim p\text{”}\,.$$

Eccone la tavola di verità:

p	q	$\sim q$	$\sim p$	$p \Longrightarrow q$	$\sim q \Longrightarrow \sim p$	PC
V	V	F	F	V	V	V
F	V	F	V	V	V	V
V	F	V	F	F	F	V
F	F	V	V	V	V	V

Se fissiamo l'attenzione sulla prima implicazione "$p \Longrightarrow q$" e la chiamiamo proposizione *diretta*, allora la seconda "$\sim q \Longrightarrow \sim p$" è detta la *contronominale*; è detta invece *inversa* la "$q \Longrightarrow p$" e *contraria* la "$\sim p \Longrightarrow \sim q$".

La tautologia PC afferma l'equivalenza logica delle proposizioni diretta e contronominale. Da essa segue la regola di deduzione (di cui si fa uso larghissimo in matematica) detta di **riduzione all'assurdo**: *nell'ipotesi che p sia vera, se si vuole dimostrare che q è vera si procede dimostrando che "$\sim q \Longrightarrow \sim p$" è vera.*

Un esempio significativo sarà dato nel par. 1.4.

Altre tautologie, la cui verifica lasciamo al lettore, sono le seguenti:

$$\begin{array}{ll}
\text{(TE)} & \text{“}p \vee \sim p\text{”} \\
\text{(SI)} & \text{“}(p \Longrightarrow q) \wedge (q \Longrightarrow r) \Longrightarrow (p \Longrightarrow r)\text{”} \\
\text{(NC)} & \text{“}\sim (p \wedge \sim p)\text{”} \\
\text{(DM)}_1 & \text{“}\sim (p \vee q) \iff \sim p \wedge \sim q\text{”} \\
\text{(DM)}_2 & \text{“}\sim (p \wedge q) \iff \sim p \vee \sim q\text{”}
\end{array}$$

La (TE) è il *principio del terzo escluso*, la (NC) è quello *di non contraddizione*. La (SI) è nota come *sillogismo ipotetico*: la prima implicazione ($p \Longrightarrow q$) è detta premessa maggiore, la seconda ($q \Longrightarrow r$) premessa minore. Le ultime due sono note come *leggi di De-Morgan*, utili quando si ha a che fare con la negazione di proposizioni composte.

È pure una tautologia la seguente: "$p \Longrightarrow q \iff \sim p \vee q$" (infatti, come abbiamo già osservato, "$p \Longrightarrow q$" e "$\sim p \vee q$" hanno la stessa tavola di verità). Dunque

anche "$\sim (p \Longrightarrow q) \Longleftrightarrow \sim (\sim p \vee q)$" è una tautologia e infine, usando DM_2, anche

$$\text{“} \sim (p \Longrightarrow q) \Longleftrightarrow p \wedge \sim q \text{”} \tag{1.1}$$

lo è: ciò significa che negare "$p \Longrightarrow q$" equivale all'enunciato: "$p \wedge \sim q$".

***Esempio* 1.1** - Siano:

p : "oggi piove"

q : "non esco di casa"

r : "guardo la partita".

"$p \Longrightarrow q$" significa: "se oggi piove non esco di casa"

"$p \wedge \sim q$" significa: "oggi piove ed esco di casa" (che è la precisa negazione dell'enunciato precedente).

Su (SI) è fondata la deduzione seguente: se oggi piove non esco di casa; se non esco di casa guardo la partita; allora, se piove guardo la partita.

$(\mathrm{DM})_1$ indica: "non è vero che: oggi piove, oppure non esco di casa, equivale ad affermare che: non piove ed esco di casa".

$(\mathrm{DM})_2$ indica: "...".

1.4 Logica dei predicati. Quantificatori

Una logica basata solo sugli enunciati è insufficiente per i nostri scopi. Occorre introdurre il concetto di *predicato*: una proposizione contenente una o più *variabili* (o argomenti). Un predicato si dice *unario* se riguarda solo una variabile e si indicherà con una scrittura del tipo $p(x)$ (si legge "p di x"). Ad esempio

$p(x)$: " x è un uomo"

$p(x)$: " x è un numero reale ≥ 1"

sono predicati unari. Analogamente, un predicato si dirà *binario, ternario*, ecc. se riguarda 2, 3 o più argomenti, come (pensando che le variabili siano numeri reali):

$p(x, y)$: "$x \geq y$"

$p(x, y, z)$: "$\sin(x + y) = z$" .

Se fissiamo tutte le variabili, il predicato diventa una proposizione che può essere vera o falsa. Ad esempio, sostituendo nel predicato: "$x \geq y$" ad x il numero 2 e ad y il numero 3, si ottiene l'enunciato falso: "$2 \geq 3$".

A partire da predicati noti si possono costruire nuovi predicati usando le parentesi ed i connettivi logici. Un'altra maniera per produrre nuovi predicati è l'applicazione dei cosiddetti quantificatori: universale ed esistenziale.

Il *quantificatore universale* è indicato col simbolo $\forall$ e significa: *per ogni*.

Il *quantificatore esistenziale* è indicato col simbolo $\exists$ e significa: *esiste*.

Per esempio, se $p(x)$ è un predicato (unario) allora:

"$\forall x: p(x)$" e "$\exists x: p(x)$"

sono predicati che significano, rispettivamente:

"per ogni x è vera $p(x)$",
"esiste almeno un x per il quale è vera $p(x)$".

Nei predicati contenenti più variabili si possono impiegare più quantificatori; bisogna allora fare attenzione all'ordine in cui questi vengono applicati, poiché il significato della proposizione risultante può essere completamente diverso nei diversi casi.

***Esempio* 1.2** - Sia $p(x,y)$: "un uomo x osserva la stella y".

$\exists x: p(x,y)$ significa: "esiste un uomo che osserva la stella y"
$\forall x: p(x,y)$ significa: "tutti gli uomini osservano y"
$\forall y, \exists x: p(x,y)$ significa: "per ogni stella esiste un uomo che la osserva"
$\exists x, \forall y: p(x,y)$ significa: "esiste un uomo che osserva ogni stella"
$\exists y, \forall x: p(x,y)$ significa: "esiste una stella osservata da tutti gli uomini"
$\forall x, \exists y: p(x,y)$ significa: "...".

Se una variabile in un predicato non è quantificata (cioè soggetta all'azione di un quantificatore) si dice *libera*. Così, considerando i vari predicati dell'esempio precedente, nei primi due la variabile y è libera; negli altri entrambe le variabili x e y sono quantificate; il predicato non dipende più effettivamente dalle variabili x e y, che pertanto si dicono *mute*. Il valore di verità di un predicato dipende dalle sue variabili libere. Se un predicato non ha variabili libere è un enunciato.

Osserviamo una importante relazione tra i due quantificatori $\exists$ e $\forall$. Si consideri, ad esempio, il predicato: "non tutte le mucche sono bianche". Se vogliamo formalizzarlo, possiamo introdurre i due predicati semplici:

$p(x)$: "x è una mucca"
$q(x)$: "x è bianco" .

Allora il nostro predicato può formalizzarsi nel modo seguente:

"$\sim (\forall x: p(x) \Longrightarrow q(x))$" .

Esso è equivalente a dire: "esiste una mucca non bianca".
Quest'ultimo enunciato può formalizzarsi così:

$$\text{"}\exists x: p(x) \wedge (\sim q(x))\text{" ovvero "}\exists x: \sim (p(x) \Longrightarrow q(x))\text{"} \tag{1.2}$$

poiché, come abbiamo già osservato (ricorda la (1.1)), la proposizione "$p \wedge \sim q$" è equivalente alla negazione della proposizione "$p \Longrightarrow q$". Siamo dunque condotti a

formulare la seguente proposizione, che, come il lettore può facilmente verificare, è sempre vera: se $a(x)$ è un predicato, in cui x è variabile libera, allora

$$\text{“} \sim [\forall x\colon a(x)] \iff [\exists x\colon \sim a(x)]\text{”} . \tag{1.3}$$

(Nell'esempio era $a(x)$: “$p(x) \implies q(x)$”). Analogamente abbiamo

$$\text{“} \sim [\exists x\colon a(x)] \iff [\forall x\colon \sim a(x)]\text{”} \tag{1.4}$$

Le (1.3), (1.4) indicano come operare col simbolo di negazione e i quantificatori: la negazione può essere scambiata con un quantificatore, pur di mutare questo nell'altro, negando il predicato quantificato.

L'equivalenza espressa in (1.1) e la regola (1.3) mettono in luce il ruolo dei *controesempi* nel dimostrare che un teorema è falso.

Si voglia ad esempio mostrare che un teorema del tipo

$$\text{“}\forall x\colon p(x) \implies q(x)\text{”}$$

è falso e cioè che “$\sim [\forall x\colon p(x) \implies q(x)]$” è vero.

In base a (1.2) e (1.3) è sufficiente esibire un particolare x (il controesempio) tale che $p(x)$ sia vera *e* $q(x)$ sia falsa.

Per esempio, per mostrare che l'enunciato: “ogni equazione di secondo grado ha una radice reale” è falso, basterà considerare la sola equazione: $x^2 + 1 = 0$.

Invece la celebre affermazione di Fermat: “non è possibile dividere un cubo in somma di due cubi, né, in generale, dividere alcun'altra potenza di grado superiore al secondo in somma di due altre potenze dello stesso grado” è una di quelle proposizioni per cui ancora non si conosce né una dimostrazione né un controesempio.

Concludiamo con un esempio di dimostrazione per assurdo: vogliamo mostrare che non esiste alcun numero razionale m/n (con m, n interi positivi primi tra loro) il cui quadrato sia uguale a 2.

Si tratta della celebre argomentazione con la quale i Pitagorici scoprirono con costernazione che la diagonale del quadrato è *incommensurabile* con il lato.

Siano:

$p(m,n)$: “m, n sono interi positivi, primi tra loro”.

$q(m,n)$: “$\dfrac{m^2}{n^2} \neq 2$”.

Si vuole provare la verità dell'enunciato

$$\text{“}\forall m, n\colon p(m,n) \implies q(m,n)\text{”} .$$

In base al principio di contrapposizione (cioè “$p \implies q \iff \sim q \implies \sim p$”) ciò equivale a dimostrare

$$\text{“}\forall m, n\colon \sim q(m,n) \implies \sim p(m,n)\text{”}$$

e cioè che se m, n sono interi positivi, tali che $m^2/n^2 = 2$ allora m ed n devono avere un fattore comune. Infatti, $m^2 = 2n^2$ implica che m^2 è pari e perciò m è

pari; dunque m^2 è divisibile per 4. Essendo $n^2 = m^2/2$, n^2 è pari. Ne segue che n è pari come m e perciò m ed n hanno il fattore 2 in comune.

Il contenuto di questa sezione potrà apparire (a qualche studente) come la complicazione inutile di regole talmente semplici da poter essere ritenute ovvie. Forse. Ma l'esperienza ci ha mostrato quante volte lo studente, anche preparato, incorra in errori di logica nel corso di una dimostrazione, non sappia formulare correttamente la negazione di un dato enunciato e così via. Perciò riteniamo che non sia una perdita di tempo l'esercizio negli argomenti sopra esposti.

Esercizi

1. Siano:

p : "Emilio ha la febbre"

q : "Emilio ha paura dell'esame"

r : "Emilio va a scuola" .

Interpretare gli enunciati "$p \vee q \Longrightarrow \sim r$" e "$\sim p \wedge \sim r \Longrightarrow q$".

2. Verificare se i seguenti enunciati sono tautologie

a) "$(p \vee q) \wedge \sim q \Longrightarrow p$"

b) "$(p \Longrightarrow q) \Longleftrightarrow (\sim p \Longrightarrow \sim q)$"

c) "$p \wedge \sim (p \vee q)$".

3. Verificare, in base al Modus-ponens, se le seguenti deduzioni sono corrette.

a) Se oggi piove esco in macchina; oggi non piove; allora non esco in macchina.

b) Se oggi piove esco in macchina; non esco in macchina; allora oggi non piove.

4. Da p e "$p \Longrightarrow q$" si deduce q poiché l'enunciato "$p \wedge (p \Longrightarrow q) \Longrightarrow q$" è una tautologia.

Generalizzando, si può dedurre q dagli enunciati p_1, $p_2, \ldots, p_n$ se l'enunciato "$p_1 \wedge p_2 \wedge \ldots \wedge p_n \Longrightarrow q$" è una tautologia.

Si esamini ora la seguente argomentazione.

Se Tizio mente allora è l'assassino oppure il delitto è avvenuto dopo la mezzanotte.

Se il delitto è avvenuto prima di mezzanotte allora Tizio mente.

Allora Tizio è l'assassino.

È corretta questa deduzione?

5. Consideriamo un circuito elettrico in cui compaiono solo interruttori a 2 posizioni: chiuso (passa corrente) ed aperto (non passa corrente). Ad ogni interruttore associamo un enunciato con la convenzione che l'interruttore è *chiuso* se e solo se l'enunciato è *vero*.

Così, nel circuito seguente,

Fig. 1.1

il passaggio di corrente è equivalente alla verità dell'enunciato "$p \wedge q$".

Si esprima la condizione di passaggio della corrente nei circuiti seguenti

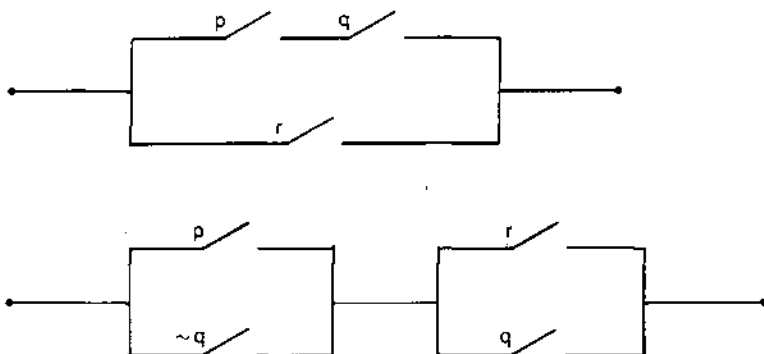



Fig. 1.2

6. Costruire un circuito del tipo indicato nell'esercizio 5 per gli enunciati

a) $p \wedge (r \vee \sim q)$, b) $p \Longrightarrow q$.

7. Siano x e y numeri naturali e $p(x,y)$: "$x < y$".
Interpretare

$$\exists x \forall y : p(x,y) \ , \ \exists x \exists y : p(x,y) \ , \ \forall x \forall y : p(x,y) \ , \ \forall x \exists y : p(x,y) \ .$$

8. Siano: $p(x)$: "x è cinese" e $q(x)$: "x è un buon cuoco".
Interpretare gli enunciati

"$\forall x : p(x) \Longrightarrow q(x)$" , "$\forall x : q(x) \Longrightarrow p(x)$"

e le loro negazioni.

9. Si consideri l'enunciato "$p \Longrightarrow q$" seguente: "se tutti gli uomini fossero sani, tutti gli ospedali sarebbero vuoti".
Scrivere la contronominale, la contraria, l'inversa di "$p \Longrightarrow q$".

2. SIMBOLI ED OPERAZIONI INSIEMISTICHE FONDAMENTALI

2.1 Definizioni

Frasi come: l'insieme dei numeri reali, l'insieme dei cittadini italiani, l'insieme dei punti di un piano, etc. sono intuitive; esse illustrano il concetto di **insieme**. In questa esposizione noi assumeremo come primitiva (cioè non riducibile a concetti più elementari) la nozione di insieme; useremo, talvolta, come sinonimi, le espressioni: *collezione, classe, famiglia, aggregato*. Indicheremo gli insiemi solitamente con le lettere maiuscole: $A, B, X, \ldots$ e con lettere minuscole $a, b, x, \ldots$ i singoli oggetti, gli elementi, che fanno parte di questi insiemi. Talvolta è utile elencare espressamente gli elementi di un insieme; ciò si fa indicando tali elementi

fra parentesi graffe. Ad esempio $\{0,1\}$ indica l'insieme formato dai numeri 0 e 1; $\{1, 1/2, 1/3, \ldots, 1/100\}$ indica l'insieme i cui elementi sono i reciproci dei numeri naturali compresi tra 1 e 100; $\{a\}$ indica l'insieme formato dal solo elemento a.

Si noti che ogni elemento di un qualsiasi insieme può, a sua volta, essere considerato un insieme; ciò è comodo da un punto di vista concettuale, poiché in tal modo nella teoria compaiono solo variabili di uno stesso tipo: gli insiemi, appunto.

Oltre ai simboli logici già introdotti nel capitolo precedente incontriamo nella teoria i simboli di *uguaglianza*: "$=$", e di *appartenenza*: "$\in$". Essi danno luogo ai predicati binari

$$x = y \quad , \quad x \in X$$

che si leggono rispettivamente: "*x è uguale a y*", "*x appartiene a X*" (oppure, "*x è un elemento di X*"). Il primo è regolato dalle proprietà:

riflessiva	$\forall x : x = x$
simmetrica	$\forall x, \forall y : (x = y) \Longleftrightarrow (y = x)$
transitiva	$\forall x, \forall y, \forall z : (x = y) \wedge (y = z) \Longrightarrow (x = z)$

e dalla seguente proprietà *di sostituzione*: se p è un predicato qualsiasi in una variabile, allora

$$\forall x, \forall y : (x = y) \Longrightarrow (p(x) \Longleftrightarrow p(y))$$

cioè x e y, se sono uguali, sono intercambiabili in ogni predicato.

Per negare i due predicati sopra introdotti si usano le comode abbreviazioni: $x \neq y$ (si legge: "*x è diverso da y*") e $x \notin X$ (si legga: "*x non appartiene a X*").

Un'altra comoda abbreviazione che useremo è il simbolo "$:=$" che viene utilizzato per porre definizioni, come per esempio: $A := \{0,1\}$ oppure $\pi :=$ (lunghezza della semicirconferenza)/(lunghezza del raggio) (*).

Risulta poi utile, per gli sviluppi formali della teoria, considerare, tra i vari insiemi, anche un insieme privo di elementi: l'*insieme vuoto*, che si indicherà col simbolo $\emptyset$.

Introduciamo ora i simboli di inclusione: $\subseteq, \supseteq, \subset, \supset$, che danno luogo anch'essi a predicati binari. Anzitutto diremo che *i due insiemi A e B sono uguali se e solo se hanno gli stessi elementi*, cioè

$$A = B \Longleftrightarrow (x \in A \Longleftrightarrow x \in B) .$$

Si dice che B è un **sottoinsieme** di A (oppure B è *contenuto* in A) e si scrive $B \subseteq A$ (oppure $A \supseteq B$), se ogni elemento di B è un elemento di A, cioè

$$\forall x : x \in B \Longrightarrow x \in A .$$

(*) Invece di $:=$ altri testi usano $\equiv$, $\stackrel{\text{def}}{=}$ o $\triangleq$.

Si noti che $A = B$ se e solo se $B \subseteq A$ e $A \subseteq B$; inoltre $\emptyset$ è contenuto in A (qualunque sia l'insieme A). Un sottoinsieme di A distinto da $\emptyset$ e da A stesso si dice *sottoinsieme proprio*; se B è un sottoinsieme proprio di A scriveremo $B \subset A$ (oppure $A \supset B$).

***Esempio* 2.1** - Se $\mathbb{R}$ indica l'insieme dei numeri reali, l'espressione

$$\{x : (x \in \mathbb{R}) \wedge (x^2 > 0)\}$$

indica il sottoinsieme proprio ottenuto da $\mathbb{R}$ togliendo lo zero.

Le relazioni di inclusione $\subseteq, \supseteq$ godono delle seguenti proprietà di verifica immediata:

riflessiva	$\forall A : A \subseteq A$
antisimmetrica	$\forall A, \forall B : (A \subseteq B) \wedge (B \subseteq A) \Longrightarrow A = B$
transitiva	$\forall A, \forall B, \forall C : (A \subseteq B) \wedge (B \subseteq C) \Longrightarrow A \subseteq C.$

Definiamo ora alcune fondamentali operazioni insiemistiche.

Nella figura 1.3, esse sono illustrate in maniera intuitiva, per mezzo dei cosiddetti *diagrammi di Venn*.

a) inclusione: $B \subset A$ *b)* unione: $A \cup B$ *c)* intersezione: $A \cap B$

d) differenza: $A \setminus B$ *e)* differenza simmetrica: $A \triangle B$ *f)* complementazione: $C_U A$

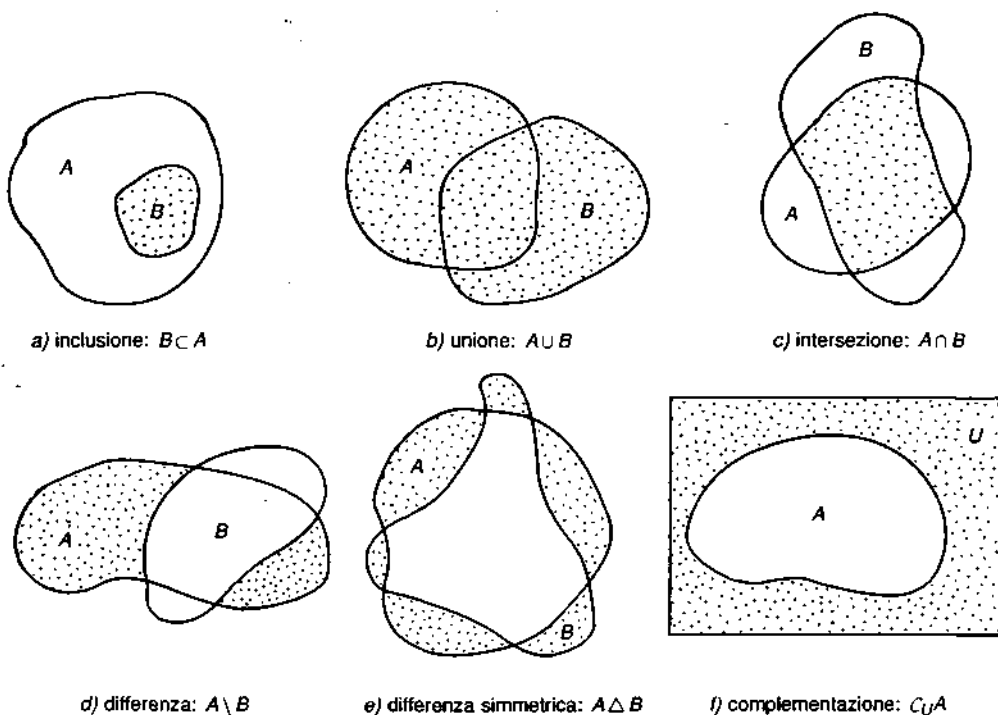



Fig. 1.3

Definizioni

2.1 Unione. *L'unione di due insiemi A, B è un insieme, indicato con $A \cup B$, costituito da tutti quegli elementi che appartengono ad almeno uno dei due insiemi A oppure B.*

$$A \cup B := \{x : (x \in A) \vee (x \in B)\} .$$

2.2 Intersezione. *L'intersezione di due insiemi A, B è un insieme indicato con $A \cap B$, costituito dagli elementi che appartengono sia ad A che a B.*

$$A \cap B := \{x : (x \in A) \wedge (x \in B)\} .$$

Se accade che $A \cap B = \emptyset$, cioè i due insiemi non hanno elementi in comune, essi si diranno *disgiunti*.

2.3 Differenza. *La differenza di B da A è un insieme, indicato con $A \setminus B$, costituito dagli elementi di A che non appartengono a B.*

$$A \setminus B := \{x : (x \in A) \wedge (x \notin B)\} .$$

È utile anche considerare la **differenza simmetrica** di A e B: essa è un *insieme*, indicato con $A \triangle B$, *costituito dagli elementi di A che non appartengono a B e dagli elementi di B che non appartengono ad A.*

$$A \triangle B := (A \setminus B) \cup (B \setminus A) .$$

2.4 Complementazione. Accade spesso, nel corso di un'argomentazione, che tutti gli insiemi di cui trattasi siano sottoinsiemi di un determinato insieme U, detto insieme ambiente.

La famiglia di tutti questi sottoinsiemi di U, propri ed impropri, prende il nome di *insieme delle parti* di U, e si indica con $\mathcal{P}(U)$. Se $A \in \mathcal{P}(U)$, si definisce allora il *complementare* di A rispetto ad U, e si indica con $C_U A$ (o semplicemente CA se l'ambiente è chiaro dal contesto) l'insieme formato dagli elementi di U che non appartengono ad A, ovvero $U \setminus A$:

$$C_U A := \{x : (x \in U) \wedge (x \notin A)\} = U \setminus A .$$

Le operazioni sopra definite godono di proprietà, le più importanti delle quali elenchiamo qui sotto, lasciandone la semplice verifica al lettore.

Sia U un dato insieme e $A, B, C, \ldots$ elementi di $\mathcal{P}(U)$; allora abbiamo:

Per l'unione	Proprietà	Per l'intersezione
$A \cup A = A$	iterativa o di idempotenza	$A \cap A = A$
$A \cup B = B \cup A$	commutativa	$A \cap B = B \cap A$
$(A \cup B) \cup C = A \cup (B \cup C)$	associativa	$(A \cap B) \cap C = A \cap (B \cap C)$
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$	distributiva	$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
$A \cup (A \cap B) = A$	di assorbimento	$A \cap (A \cup B) = A$
$A \subseteq B \iff (A \cup B = B)$	di inclusione	$A \supseteq B \iff (A \cap B = B)$

Per la complementazione	Proprietà
$\mathcal{C}(\mathcal{C}A) = A$	involutoria
$\mathcal{C}(A \cup B) = \mathcal{C}A \cap \mathcal{C}B$ $\mathcal{C}(A \cap B) = \mathcal{C}A \cup \mathcal{C}B$	pseudodistributive, o leggi di De Morgan (cfr. le tautologie DM_1 e DM_2 del par 1.1)
$A \cup \mathcal{C}A = U, A \cap \mathcal{C}A = \emptyset$	

L'insieme $\mathcal{P}(U)$, sugli elementi del quale sono state definite le operazioni di unione, intersezione, complementazione come sopra esposto (le altre operazioni possono esprimersi per mezzo di queste) costituisce un'**algebra di Boole** (fondata sull'insieme U). Si noterà come le precedenti proprietà si presentino a coppie; questo fatto esprime la **legge di dualità**: *se, in una proprietà vera, espressa mediante i simboli di inclusione, di unione e intersezione, si scambiano tra loro unione e intersezione e ogni inclusione* $\supseteq$ *con la sua opposta* $\subseteq$, *il risultato è una proposizione ancora vera; da una proprietà vera, espressa mediante i simboli di* $\cup$, $\cap$, $\mathcal{C}$, *si ottiene una proprietà ancora vera scambiando fra loro* $\cup$ *e* $\cap$ *e ogni insieme col suo complementare.*

Vogliamo concludere questa introduzione mostrando che, per quanto intuitivi e semplici possano sembrare i concetti sopra esposti, essi ci portano immancabilmente incontro a delle contraddizioni, note come *paradossi* della teoria degli insiemi. Infatti noi abbiamo tacitamente assunto che ogni predicato possa *caratterizzare* l'appartenenza ad un insieme.

In altri termini, se $p(x)$ è un predicato unario, risulta definito l'insieme $X := \{x : p(x)\}$. Viceversa, dato un insieme X è definito il predicato $p(x)$: "$x \in X$".

Le definizioni di $A \cup B$, $A \cap B$, $A \setminus B$ etc, date sopra, sono state ottenute in questo modo. Ora, a parte il fatto che il predicato $p(x)$ dovrà soddisfare evidenti requisiti

di sensatezza (non ha senso dire: l'insieme dei numeri simpatici!) succede che il semplice predicato:

$p(x)$: "x è un insieme"

porta ad una contraddizione.

Infatti, l'insieme $I := \{x: p(x)\}$ sarebbe l'*insieme di tutti gli insiemi*. Di conseguenza $I \in I$ e cioè I apparterrebbe a sé stesso.

Dividiamo gli insiemi in due categorie: la categoria R (R da Russell), costituita dagli insiemi che appartengono a sé stessi; un esempio di insieme di tipo R è proprio I, l'insieme di tutti gli insiemi. La categoria $\sim R$ (leggiamo "non R"), costituita da tutti gli altri insiemi.

Naturalmente, gli insiemi più comuni sono di tipo $\sim R$; ad esempio $\mathbb{N}$, l'insieme dei numeri naturali, *non* è un numero naturale; un insieme di dieci mele *non* è una mela.

Consideriamo ora l'insieme definito da

$$H := \{X: X \text{ è un insieme di tipo } \sim R\}\ .$$

In altri termini H è l'insieme di tutti gli insiemi di tipo $\sim R$.

Per il principio del terzo escluso, "$p \vee (\sim p)$", abbiamo che H è di tipo R oppure H è di tipo $\sim R$.

Tuttavia osserviamo che:
se H fosse di tipo R, allora si avrebbe $H \in H$; ma *ogni* elemento di H è un insieme di tipo $\sim R$ e si dedurrebbe così che H è di tipo $\sim R$. Contraddizione.

Supponiamo allora che H sia di tipo $\sim R$; allora, per la definizione di H, si avrebbe $H \in H$ con la conseguenza che H è di tipo R. Ancora contraddizione.

Conclusione: o il principio del terzo escluso (che è una tautologia!) non vale oppure la teoria è contraddittoria.

Quello che abbiamo illustrato è il celebre paradosso di Bertrand Russell. Un altro paradosso, basato sul concetto di *cardinalità*, dovuto allo stesso Cantor (George Cantor è il fondatore della teoria degli insiemi) è presentato più avanti.

Ciò che, in ultima analisi, risulta contraddittorio è il concetto di insieme di tutti gli insiemi oppure di insieme che appartiene a sé stesso.

La necessità di impedire l'insorgere di tali paradossi ha portato alla formulazione assiomatica della teoria degli insiemi. Per soddisfare la curiosità di un eventuale lettore interessato ed anche perché avremo occasione di citare più avanti qualcuno degli assiomi della teoria, ne abbiamo inserito un rapido cenno nell'appendice in fondo al capitolo.

Per ovvie ragioni di semplicità, continueremo ad usare la teoria non assiomatica o, come si usa dire, "ingenua"; avremo però cura di evitare riferimenti all'"insieme di tutti gli insiemi" o a predicati del tipo $X \in X$.

Tale modo di procedere è parzialmente giustificato dal fatto che, come recenti sviluppi di logica, sui quali certo non possiamo soffermarci, hanno dimostrato, anche la formulazione assiomatica non è esente da critiche.

In ogni caso si può affermare che la teoria degli insiemi costituisce una delle rivoluzioni più *profonde* nella storia della scienza.

Esercizi

1. Sia A un insieme ed $a \in A$.
È vero che $A \cup \{a\} = A$? È vero che $A \cup \{\{a\}\} = A$?

2. Siano A, B, C parti di un dato insieme ambiente U; stabilire quali delle seguenti relazioni sono vere:

a) $A \subseteq C$ e $B \subset C \Longrightarrow A \cup B \subset C$

b) $A \subseteq C$ e $A \subseteq B \Longrightarrow A \subseteq B \cap C$

c) $A \subseteq B \Longrightarrow \complement A \subseteq \complement B$.

3. Dimostrare che

$$A \cup B = (A \cap B) \cup (B \setminus A) \cup (A \setminus B) .$$

4. Dimostrare le leggi di De Morgan

$$\complement(A \cup B) = \complement A \cap \complement B$$
$$\complement(A \cap B) = \complement A \cup \complement B .$$

5. Sia B l'insieme dei barbieri di Lodi che radono la barba a quelli e soltanto a quelli che non se la radono da soli.
Dimostrare che o $B = \emptyset$, oppure i barbieri appartenenti a B hanno barbe ... chilometriche.

6. Quale delle seguenti relazioni è vera?

a) $A \supset B$ e $C \cap A \neq \emptyset \Longrightarrow C \cap B = \emptyset$

b) $A \supset B \cap C$ e $A \cap B = \emptyset \Longrightarrow B \cap C = \emptyset$

c) $(A \cap B) \setminus C = (A \cap C) \setminus (B \cap C)$.

3. RELAZIONI

3.1 Prodotto cartesiano

In maniera intuitiva introduciamo la nozione di **coppia ordinata**. Presi due insiemi A e B (non necessariamente distinti) indichiamo con (a, b) l'insieme costituito prendendo due elementi $a \in A$ e $b \in B$ nell'ordine indicato. Sia (a', b') un'altra coppia, $a' \in A$, $b' \in B$. Definiamo l'uguaglianza tra coppie ordinate ponendo

$$(a, b) = (a', b') \Longleftrightarrow (a = a') \wedge (b = b')$$

Perciò è chiaro che (a, b) non va confusa con la coppia (non ordinata) $\{a, b\}$, cioè l'insieme dei due elementi $a \in A$, $b \in B$, che può essere anche indicato con $\{b, a\}$ poiché non importa l'ordine con cui si elencano gli elementi di un insieme.

Definizione 3.1 - *La collezione di tutte le coppie ordinate (a,b) ottenute prendendo $a \in A$ e $b \in B$ forma un nuovo insieme che si indica con $A \times B$ e si chiama* **prodotto cartesiano** *(o semplicemente prodotto) di A per B:*

$$A \times B := \{(a,b): a \in A, b \in B\} .$$

È chiaro che, se A è diverso da B, questo prodotto non è commutativo e cioè $A \times B \neq B \times A$; nel caso $A = B$ esso si indicherà semplicemente con A^2.

***Esempio* 3.1** - Se A e B sono sottoinsiemi di $\mathbb{R}$ (numeri reali), le coppie (a,b) si possono rappresentare come punti del piano, convenendo che il primo elemento della coppia sia l'ascissa e il secondo l'ordinata. Allora, sia A l'insieme dei numeri compresi tra 0 e 1 e B l'insieme dei numeri compresi tra 1 e 2; la figura 1.4 illustra gli insiemi $A \times B$, $B \times A$, A^2.

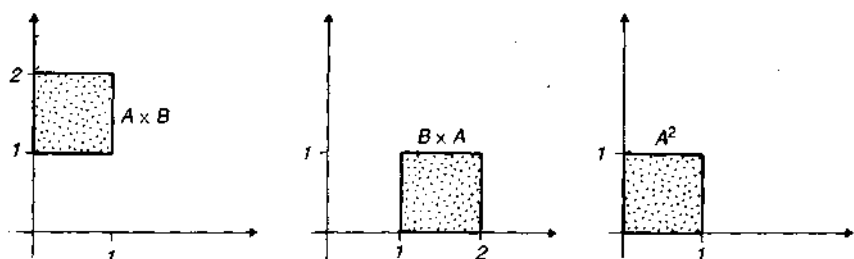



Fig. 1.4

La definizione si estende poi in maniera ovvia al caso di più di due fattori: $A_1 \times A_2 \times \ldots \times A_n$ è l'insieme delle n-uple ordinate $(a_1, a_2, \ldots, a_n)$ con $a_i \in A_i$ $(i = 1, 2, \ldots, n)$.

Lo studente mostri che, per questo prodotto, valgono la proprietà associativa e la proprietà distributiva rispetto all'unione, intersezione e differenza.

3.2 Definizione di relazione

Definizione 3.2 - *Siano X e Y due insiemi e per $x \in X$, $y \in Y$, sia $r(x,y)$ un predicato (binario) nelle variabili x e y. Si dice che questo predicato esprime una* **relazione** *binaria (o corrispondenza) tra gli elementi di X e di Y. Se $r(x,y)$ è vera si dice che x e y stanno in relazione tra loro.*

Se $X = Y$, diremo semplicemente che r è una relazione tra gli elementi di X.

Sia R il sottoinsieme di $X \times Y$ costituito dalle coppie (x,y) per cui la relazione è vera:

$$R := \{(x,y): (x,y) \in X \times Y \wedge r(x,y)\} .$$

R si chiama *grafico* della relazione e si indica con $\text{graf}(r)$. Per indicare che x e y sono legati dalla relazione r scriveremo

$$(x, y) \in \text{graf}(r) .$$

Viceversa, dato un insieme $R \subseteq X \times Y$, risulta individuata la relazione r definita da:

"$r(x, y)$ vera se e solo se $(x, y) \in R$" .

Evidentemente il grafico di questa relazione coincide con R.

In ultima analisi, una relazione binaria r in $X \times Y$ è identificabile con un sottoinsieme di $X \times Y$, il grafico di r.

Nella lingua parlata: "x è fratello di y", "x abita nella città y", "x è iscritto alla facoltà y", etc., sono esempi evidenti di relazioni. A noi interessano le relazioni espresse nel linguaggio della teoria degli insiemi.

Esempi

3.2. X è l'insieme dei punti del piano, Y l'insieme delle rette del piano: una relazione in $X \times Y$ è quella di appartenenza del punto alla retta.

3.3. Dato un insieme U, sia $X = Y = \mathcal{P}(U)$. Allora se $A, B \in \mathcal{P}(U)$ il predicato $r(A, B)$: "$A \subseteq B$" esprime una relazione in $\mathcal{P}(U)$.

Se $U = \{a, b\}$ lo studente rappresenti schematicamente con un disegno il grafico della relazione.

3.4. Sia X l'insieme degli interi positivi. Esempi di relazioni in X sono:

$r(m, n)$: "m ed n sono primi tra loro";

$s(m, n)$: "m è divisore di n";

$t(m, n)$: "$m^2 + n^2 =$ quadrato di un intero".

Lo studente cerchi di farsi un'idea del grafico di queste relazioni, magari limitandosi a considerare gli interi compresi tra 1 e 100.

3.5. Sia X l'insieme dei numeri reali. Esempi di relazioni in X sono: "$x \leq y$", "$x^2 + y^2 + 1 = 0$", "$(\sin x)^2 + (\cos y)^2 = 1$".

Qual è il loro grafico? (Quando il grafico è vuoto, la relazione vien detta *impossibile*).

Sono di grande importanza due speciali tipi di relazioni definite fra gli elementi di uno stesso insieme: le equivalenze e gli **ordinamenti**.

3.3 Equivalenze

Definizione 3.3 - *Diciamo che r è una* **relazione di equivalenza** *(o semplicemente un'equivalenza) in X, e la indicheremo col simbolo $x \approx y$ (si legga: x è*

equivalente a y) se verifica le proprietà:

riflessiva $\forall x \in X : x \approx x$
simmetrica $\forall x \in X, \forall y \in X : x \approx y \Longleftrightarrow y \approx x$
transitiva $\forall x \in X, \forall y \in X, \forall z \in X : (x \approx y) \wedge (y \approx z) \Longrightarrow x \approx z$.

L'importanza di una simile relazione sta nel fatto che essa realizza, sull'insieme in cui è definita, una **partizione**. Definiamo cosa si intende con questo.

Definizione 3.4 - *Sia X un insieme non vuoto. Una* **partizione** *di X, che indicheremo con S, è una famiglia di sottoinsiemi di X, ovvero un sottoinsieme di $\mathcal{P}(X)$ tale che:*

i) *ogni elemento A di S è non vuoto*
ii) *se $A_1 \in S$, $A_2 \in S$ e $A_1 \neq A_2$, allora $A_1 \cap A_2 = \emptyset$*
iii) $\bigcup\limits_{A \in S} A = X$.

Per esempio, se X è l'insieme degli interi positivi, una sua partizione si ha prendendo $S = \{P, D\}$ dove P è l'insieme dei numeri pari e D l'insieme di quelli dispari.

Se in X è data una equivalenza r, indichiamo con $[x]$ l'insieme di tutti gli elementi di X equivalenti a x, cioè

$$[x] := \{y : (y \in X) \wedge (x \approx y)\} ;$$

$[x]$ è detta **classe di equivalenza** di x.

■ **Teorema 3.1** - *Stabilita in un insieme non vuoto X una relazione di equivalenza, la famiglia $\{[x] : x \in X\}$ delle classi di equivalenza costituisce una partizione di X.*

Dimostrazione - Dobbiamo mostrare che valgono le proprietà i) ii) iii) sopra menzionate. La i) è evidente poiché $[x]$ contiene almeno x.

Per mostrare la ii) procediamo per assurdo: supponiamo che due classi di equivalenza distinte $[x]$ e $[y]$ abbiano un elemento in comune, z; sarebbe allora $x \approx z$ e $y \approx z$ e perciò, per la proprietà transitiva, $x \approx y$; qualunque altro elemento $x' \in [x]$, essendo equivalente a x, sarebbe perciò equivalente a y e qualunque altro elemento $y' \in [y]$, essendo equivalente a y sarebbe perciò equivalente a x, da cui seguirebbe $[x] = [y]$ contro la ipotesi.

La iii) è pure evidente, poiché un qualsiasi elemento di X appartiene alla classe che egli stesso individua. □

Possiamo così affermare che: *ogni elemento $x \in X$ sta in una e una sola classe di equivalenza.* La partizione determinata su X dall'equivalenza r prende un nome speciale: viene detta **insieme quoziente** di X rispetto a r e viene indicata col simbolo X/r.

È anche immediato constatare che, assegnata una partizione S su X, questa determina univocamente una equivalenza r, tale che $S = X/r$. Tale equivalenza è

definita da:

"$x \approx y \iff x$ e y appartengono allo stesso elemento di S" .

Esempi

3.6. La relazione di uguaglianza, già introdotta in 2.1, è evidentemente una equivalenza; in questo caso le classi $[x]$ contengono solamente l'individuo x.

3.7. In algebra e in teoria dei numeri hanno grande importanza le **classi di resti modulo** m, definite dalla seguente equivalenza.
Sia $\mathbb{Z}$ l'insieme dei numeri interi relativi $(0, \pm 1, \pm 2, \ldots)$ e sia m un intero fissato ≥ 1; consideriamo il predicato binario

$r(x, y)$: "$x - y$ è divisibile per m" .

È facile verificare (esercizio) che r è una relazione di equivalenza in $\mathbb{Z}$. Ogni classe è composta da numeri che, divisi per m, danno lo stesso resto; da qui la terminologia.
Se $m = 1$, la r definisce una sola classe che contiene tutti gli interi.
Se $m = 2$, la r definisce due classi, una formata dai numeri pari, l'altra dai dispari.
Se $m = 3$, la r definisce 3 classi: quella dei multipli di 3, quella dei (multipli di 3)+1, quella dei (multipli di 3)+2.

[0]	[1]	[2]	[3]	[4]
⋮	⋮	⋮	⋮	⋮
−15	−14	−13	−12	−11
−10	−9	−8	−7	−6
−5	−4	−3	−2	−1
0	**1**	**2**	**3**	**4**
5	6	7	8	9
10	11	12	13	14
15	16	17	18	19
⋮	⋮	⋮	⋮	⋮

Fig. 1.5 - Le classi di resti modulo 5.

3.8. Nella geometria euclidea sono relazioni di equivalenza il *parallelismo* tra le rette di un piano, la *similitudine* tra i poligoni di un piano; ricordiamo anche che proprio in geometria si dicono equivalenti due figure piane (o solide) che hanno la stessa area (o volume).

3.9. In fisica molte grandezze (quali forze, velocità, accelerazione, etc.) si esprimono per mezzo di **vettori**. Questi vengono di solito definiti col seguente procedimento.

Nello spazio sono assegnati, nell'ordine, due punti P e Q; risulta quindi individuato un *segmento orientato*, da P verso Q; esso può essere caratterizzato assegnando:

a) un numero (reale non negativo) che esprime la sua *lunghezza* o *intensità*;

b) una retta (quella congiungente P con Q) che individua la sua *direzione*;

c) un *verso* (cioè quello da P verso Q);

d) un punto (il punto P, origine del segmento orientato).

Consideriamo ora, tra tutti i segmenti orientati dello spazio, la seguente relazione, detta **equipollenza**: *due segmenti orientati s e s' sono equipollenti se uno di essi può sovrapporsi all'altro con una traslazione.*

Questa relazione è una equivalenza. Le classi di questa equivalenza sono i **vettori**, i quali perciò sono individuati solamente da *intensità*, *direzione* e *verso*.

Altre interessanti relazioni di equivalenza si incontreranno nel Cap. 2.

3.4 Ordinamenti

Definizione 3.5 - *Diremo che r è una* **relazione d'ordine** (o semplicemente un **ordinamento**) *in X, e la indicheremo col simbolo $x \preceq y$* (oppure $y \succeq x$) (si legga: *x precede y, oppure, y segue x*) *se verifica le proprietà:*

riflessiva $\quad \forall x \in X : x \preceq x$

antisimmetrica $\quad \forall x \in X, \forall y \in X : (x \preceq y) \wedge (y \preceq x) \iff x = y$

transitiva $\quad \forall x \in X, \forall y \in X, \forall z \in X :$

$$(x \preceq y) \wedge (y \preceq z) \implies x \preceq z \; .$$

Gli elementi (x, y) del grafico della relazione si diranno tra di loro *confrontabili*. Se, presi comunque x e $y \in X$, possiamo stabilire che $x \preceq y$ oppure $y \preceq x$, l'ordinamento si dirà *totale*, altrimenti verrà detto *parziale*. Un insieme in cui sia stata introdotta una relazione d'ordine si dirà *ordinato*.

Esempi

3.10. In $\mathbb{R}$ (insieme dei numeri reali) è definito un ordinamento totale per mezzo della usuale relazione $\leq, \geq$. In $\mathbb{R}$ cioè, $x \preceq y$ significa $x \leq y$.

3.11. Dato l'insieme U, in $\mathcal{P}(U)$ è definito un ordinamento parziale per mezzo della relazione di inclusione $\subseteq$ (vedi 2.1). Più precisamente se $A, B \in \mathcal{P}(U)$, $A \preceq B$ significa $A \subseteq B$.

3.12. L'insieme delle parole di un dizionario è totalmente ordinato secondo l'ordine lessicografico.

3.13. L'insieme dei *vettori* (precedentemente definiti) è ordinato (totalmente? parzialmente?) rispetto alla relazione definita da: $x \preceq y \iff$ la lunghezza di x è $\leq$ della lunghezza di y?

In un insieme ordinato possiamo introdurre concetti che svolgeranno un ruolo fondamentale nel seguito. Essi vengono ora solo brevemente accennati; la loro importanza verrà meglio compresa più avanti.

Sia X un insieme ordinato e sia $A \subseteq X$ *non vuoto*. Evidentemente anche A è ordinato, con l'ordinamento indotto da X. Un elemento k di X è un **maggiorante** (risp. **minorante**) di A se:

i) k è confrontabile con ogni elemento di A

ii) $\forall x \in A: x \preceq k$ (risp. $x \succeq k$).

Se esiste almeno un maggiorante (minorante) il sottoinsieme A viene detto *limitato superiormente (limitato inferiormente)*; si dirà *limitato* se risulta limitato sia superiormente che inferiormente. Un elemento m di X è detto *massimo* (risp. *minimo*) di A se:

i) $m \in A$

ii) m è un maggiorante (risp. minorante) di A.

Evidentemente, un insieme A può non avere alcun maggiorante né minorante, oppure può avere più maggioranti e più minoranti; *ma non può esservi che un solo massimo e un solo minimo.* Tuttavia questi elementi possono facilmente mancare, come accade, per esempio, per il sottoinsieme A di $\mathbb{R}$ formato dai numeri reali $x: 0 < x < 1$. Conviene allora estendere la definizione di massimo e minimo in modo da avere a disposizione delle quantità la cui esistenza, almeno nei casi che ci interesseranno in seguito, potrà agevolmente essere provata.

Definizione 3.6 - *Chiameremo* **estremo superiore (inferiore)** *di A, e lo indicheremo con* $\sup A$ ($\inf A$), *il minimo (massimo) dei maggioranti (minoranti) di A, se esiste.*

L'estremo superiore e l'estremo inferiore, se esistono, sono unici. È evidente infine che se m è massimo (minimo) di A esso coincide con $\sup A$ ($\inf A$).

Chiudiamo il paragrafo con un'ultima definizione. *Un insieme ordinato X è detto* **bene ordinato** *se ogni sottoinsieme, non vuoto, ordinato di X possiede un elemento minimo.*

Una conseguenza di questa definizione è che ogni insieme bene ordinato è totalmente ordinato poiché, presi x e y in X, l'insieme $\{x, y\}$ è un sottoinsieme di X e pertanto ha minimo; secondoché questo minimo sia x o y potremo scrivere $x \preceq y$ oppure $y \preceq x$.

Esercizi

1. Siano A un cerchio e B un segmento di retta. Interpretare geometricamente $A \times B$.

2. Siano A un cerchio di raggio r e B una circonferenza di raggio $R > r$. Interpretare geometricamente $A \times B$.

3. Sia X l'insieme dei numeri reali; disegnare il grafico delle seguenti relazioni in $X \times X$:

a) $9x^2 + 4y^2 = 36$ b) $y = 3x^2 + 1$.

4. Nell'insieme degli esseri umani esaminare la relazione

$r(x,y)$: "x è discendente da y",

ovvero "y è antenato di x".

È una relazione di equivalenza? È una relazione d'ordine?

5. Sia X l'insieme degli interi positivi e $s(m,n)$ la relazione in X data da: "m è divisore di n".
È una relazione d'ordine totale o parziale, in X?

6. Verificare che la relazione di congruenza tra figure piane o solide nella geometria euclidea è una relazione di equivalenza.

7. Dimostrare che ogni insieme totalmente ordinato e finito (cioè con un numero finito di elementi) possiede massimo e minimo.

4. FUNZIONI

4.1 Definizione di funzione

Il concetto di funzione è centrale per tutto il corso di Analisi.

Definizione 4.1 - *Siano X, Y due insiemi. Si dice* **funzione** *da X in Y una relazione r in $X \times Y$ che soddisfi la seguente proprietà:*
per ogni $x \in X$, esiste uno ed un solo elemento $y \in Y$ in relazione con x. Ovvero, l'insieme

$$\{y \in Y : (x,y) \in \text{graf}(r)\}$$

è costituito da un solo elemento per ogni $x \in X$.

La proprietà caratterizzante le funzioni si può esprimere nella maniera alternativa seguente:

due coppie (x, y_1) e (x, y_2) appartenenti al grafico di r che hanno lo stesso primo elemento devono avere uguale anche il secondo; ovvero:

$$(x, y_1) \in \text{graf}(r) \quad \text{e} \quad (x, y_2) \in \text{graf}(r) \Longrightarrow y_1 = y_2 \; .$$

***Esempio* 4.1** - Consideriamo i due casi seguenti di relazioni in $X \times Y$ dove X ed Y sono sottoinsiemi di $\mathbb{R}$, l'insieme dei numeri reali. In base a quanto detto

in 3.2 possiamo limitarci ad assegnare il loro grafico che nel nostro caso è un sottoinsieme di $\mathbb{R}^2$.

a) grafico della relazione r_1 *b)* grafico della relazione r_2

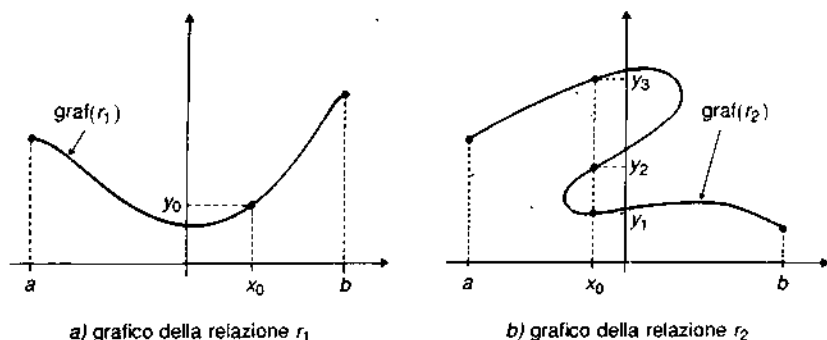



Fig. 1.6

Per entrambe le relazioni abbiamo

$$X = \{x \in \mathbb{R}: a \leq x \leq b\} .$$

Per ogni $x_0 \in X$, l'insieme $\{y \in Y: (x_0, y) \in \mathrm{graf}(r_1)\}$ è costituito esattamente da un punto: l'ordinata del punto di intersezione della retta di equazione $x = x_0$ con $\mathrm{graf}(r_1)$. Dunque r_1 è una funzione.

Al contrario l'insieme $\{y \in Y: (x_0, y) \in \mathrm{graf}(r_2)\}$ non è costituito da un solo punto per ogni $x_0 \in X$: i punti (x_0, y_1), (x_0, y_2), (x_0, y_3) indicati in figura 1.6 b) appartengono a $\mathrm{graf}(r_2)$. Dunque r_2 *non* è una funzione.

Ritorniamo alla definizione 4.1. In base ad essa ad ogni elemento $x \in X$ risulta associato un unico elemento $y \in Y$ tale che $(x, y) \in \mathrm{graf}(r)$.

Questa osservazione permette una definizione alternativa di funzione, più "dinamica" rispetto alla 4.1, preferibile dal punto di vista pratico.

Definizione 4.1$'$ - *Siano X ed Y due insiemi. Si dice* **funzione** *da X in Y una* **corrispondenza univoca** *da X in Y, ovvero una corrispondenza che associa ad ogni elemento $x \in X$ uno ed un solo elemento $y \in Y$.*

Esempi

4.2. Siano: X l'insieme degli uomini, Y l'insieme delle donne.
La corrispondenza

$$\text{uomo} \quad \mapsto \quad \text{madre}$$

è una funzione da X in Y.

4.3. Siano: X l'insieme dei padri, Y l'insieme degli uomini.
La corrispondenza

$$\text{padre} \quad \mapsto \quad \text{figlio}$$

non è una funzione da X in Y non essendo *univoca*: un padre può avere più di un figlio.

4.4. Siano: X l'insieme dei triangoli di un piano, Y l'insieme dei numeri reali. La corrispondenza

triangolo $\mapsto$ perimetro

è una funzione da X in Y.

Sia f una funzione da X in Y. X si chiama **dominio** di $f: X = \mathrm{dom}(f)$; Y si chiama **codominio** di $f: Y = \mathrm{cod}(f)$.

Se $x \in X$ e $y \in Y$ è l'elemento che f associa ad x si scrive

$y = f(x)$ (si legga "y uguale ad f di x")

e *si dice che "y è immagine di x tramite f".*

L'insieme delle immagini di tutti i punti di X tramite f si indica con $\mathrm{im}(f)$ o con $f(X)$ e si chiama **immagine di** X (tramite f). In formule:

$$f(X) := \{y \in Y : \exists\, x \in X : f(x) = y\} .$$

In alcuni testi per codominio si intende $f(X)$.

In generale $f(X) \subseteq Y$. Può capitare che $f(X)$ sia costituito da un solo elemento di Y. In tal caso la funzione si dice *costante.*

Il diagramma sottostante illustra i concetti suesposti.

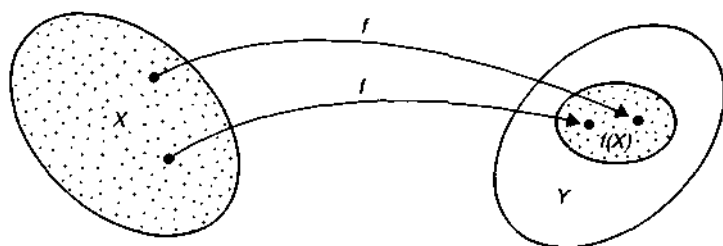



Fig. 1.7

Si può pensare ad una funzione come ad una "scatola" che trasforma un elemento del dominio inserito in INPUT in un unico elemento del codominio (OUTPUT)

INPUT $x \longrightarrow$ [f] OUTPUT $\longrightarrow f(x)$

Una notazione completa per le funzioni è la seguente:

$$f\colon\ x \mapsto f(x) : X \to Y \ .$$

Il simbolo $x \mapsto f(x)$ precisa la corrispondenza, mentre $X \to Y$ precisa dominio e codominio.

Si usa anche scrivere $f\colon X \to Y$ precisando a parte la corrispondenza: $f(x) = \ldots$.

Esempi

4.5. $f\colon x \mapsto \sin x\colon \mathbb{R} \to \mathbb{R}$
indica la corrispondenza che ad ogni numero reale x associa il seno di x. Per questa funzione si ha

$$f(\mathbb{R}) = \{x \in \mathbb{R}\colon -1 \le x \le 1\} \ .$$

4.6. $f\colon (x_1, x_2, x_3) \mapsto \sqrt{x_1^2 + x_2^2 + x_3^2} : \mathbb{R}^3 \to \mathbb{R}$
indica la corrispondenza che associa ad ogni punto di $\mathbb{R}^3$ la sua distanza dall'origine. Si ha

$$f(\mathbb{R}^3) = \{x \in \mathbb{R}\colon x \ge 0\} \ .$$

4.7. $f\colon x \mapsto (\cos x, \sin x)\colon \mathbb{R} \to \mathbb{R}^2$
indica la corrispondenza che associa ad ogni punto $x \in \mathbb{R}$ il punto P del piano sulla circonferenza con centro nell'origine indicato in figura 1.8. Per questa funzione si ha:

$$f(\mathbb{R}) = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 = 1\}$$

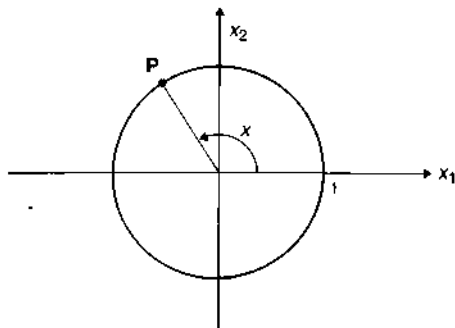



Fig. 1.8

4.8. Sia f la corrispondenza così definita:

$$f\colon \mathbb{R} \to \mathbb{R}$$

è tale che

$$f(x) = 0 \text{ se } x \text{ è irrazionale}$$
$$f(x) = 1 \text{ se } x \text{ è razionale.}$$

Sinteticamente possiamo scrivere, indicando con $\mathbb{Q}$ l'insieme dei numeri razionali

$$f(x) = \begin{cases} 1 & \text{se } x \in \mathbb{Q} \\ 0 & \text{se } x \in \mathbb{R} \setminus \mathbb{Q}. \end{cases}$$

Si ha $f(\mathbb{R}) = \{0, 1\}$.

4.9. Sia $A \subseteq \mathbb{R}^n$ ed f la corrispondenza così definita:

$$f: \mathbb{R}^n \to \mathbb{R} : f(x) = 0 \text{ se } x \notin A \ , \ f(x) = 1 \text{ se } x \in A \ .$$

Si ha $f(\mathbb{R}^n) = \{0, 1\}$.

Questa funzione si chiama *funzione caratteristica* o funzione indicatrice di A in $\mathbb{R}^n$. Si indica col simbolo $\mathbb{1}_A$.

Evidentemente la funzione dell'esempio 4.8 è l'indicatrice dei razionali in $\mathbb{R}$.

Uguaglianza tra funzioni

Nella definizione di funzione compaiono tre elementi: il dominio, il codominio, la legge che determina la corrispondenza tra essi.

Dunque due funzioni *sono uguali* se questi tre elementi coincidono per esse. Di conseguenza le funzioni

$$f: x \mapsto x^2 : \mathbb{R} \to \mathbb{R}$$

e

$$g: x \mapsto x^2 : \mathbb{R} \to \mathbb{R}_+ \cup \{0\} \qquad (\mathbb{R}_+ := \{x \in \mathbb{R}: x > 0\})$$

sarebbero da considerarsi *diverse*, avendo codominio diverso.

Dal punto di vista pratico è però conveniente "rilassare" la condizione di uguaglianza lasciando una certa libertà di scelta sul codominio, focalizzando l'attenzione solo sul dominio e sulla legge. Così facendo si potranno ritenere uguali due funzioni se hanno lo stesso grafico.

Con questa posizione le due funzioni f e g definite sopra si possono considerare uguali.

Così, nel seguito, soprattutto quando X e Y saranno insiemi numerici, si potrà definire una funzione assegnando solo l'espressione analitica che determina la corrispondenza, assumendo che il dominio sia quello "naturale", costituito cioè da tutti gli elementi per i quali l'espressione ha senso e pensare che il codominio sia Y oppure qualunque altro insieme, purché ovviamente contenga l'immagine.

Ad esempio, assegnando la corrispondenza

$$f: x \mapsto \frac{\sqrt{x}}{1 - x}$$

si assume come dominio il seguente:

$$\text{dom}(f) = \{x \in \mathbb{R} : x \geq 0 \text{ e } x \neq 1\} .$$

Quanto al codominio, si può scegliere l'insieme $\mathbb{R}$ dei numeri reali oppure qualunque insieme contenente l'immagine $f(\mathbb{R})$.

Si potrebbe pensare che le fastidiose precisazioni riguardanti il codominio possano venire eliminate scegliendo *sempre* come codominio l'insieme delle immagini, operando dunque con funzioni

$$f : X \to f(X) .$$

Se dal punto di vista teorico tale posizione è accettabile, lo è molto meno dal punto di vista pratico poiché è spesso complicato determinare l'insieme delle immagini.

Ci concederemo quindi la libertà di scelta sul codominio menzionata sopra, nel caso in cui dovremo confrontare o definire funzioni.

Una certa cautela nella precisazione del codominio sarà richiesta nel problema dell'invertibilità di una funzione, trattato più avanti, nel par. 4.4.

Alla fine del volume sono riportati i diagrammi (grafici) di numerose funzioni reali di variabile reale (cioè $f : X \to Y$ con $X \subseteq \mathbb{R}$ e $Y \subseteq \mathbb{R}$) alcune delle quali sono certamente già note allo studente dalle scuole inferiori. È utile comunque che l'allievo prenda subito conoscenza di alcune di esse, così da aver sottomano un adeguato numero di esempi sui quali appoggiare i concetti generali già introdotti o che introdurremo.

4.2 Funzioni particolari. Successioni. Multifunzioni

Un po' di definizioni, di frequente uso nel seguito.

La funzione $x \mapsto x : X \to X$ si dice *applicazione identica* (o *identità*) di X in sé: si indica con I_X.

Sia assegnata $f : X \to Y$ e sia Z un sottoinsieme di X; risulta definita in maniera naturale una funzione $g : Z \to Y$ in questo modo: $g(x)$ si pone uguale a $f(x)$ per $x \in Z$. Questa funzione g si chiama la *restrizione* di f a Z (f si dirà una *estensione* di g a X) e si indica con $f_{|Z}$.

La restrizione della funzione identica I_X ad un sottoinsieme Z di X si chiama *inclusione* di Z in X.

Dati due insiemi X e Y, risultano definite, in maniera semplice e naturale, le due funzioni seguenti:

$$f : (x, y) \mapsto x : X \times Y \to X \quad \text{e} \quad g : (x, y) \mapsto y : X \times Y \to Y .$$

Queste funzioni si chiamano, rispettivamente, *proiezione canonica* su X, *proiezione canonica* su Y.

Sia $f: X \to Y$ e sia A un sottoinsieme di X; si dice *immagine* di A e si indica con $f(A)$, il sottoinsieme B di Y costituito dagli elementi che provengono da qualche elemento di A, cioè:

$$f(A) := \{y \in Y : \ \exists \ x \in A \ : \ f(x) = y\} .$$

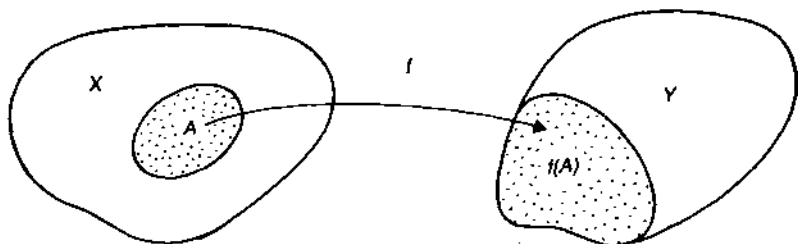



Fig. 1.9

Successioni - Applicazioni di particolare importanza sono quelle definite su $\mathbb{N}$ (l'insieme dei numeri naturali). Esse hanno un nome ed una notazione tradizionali: si chiamano *successioni* e invece di scrivere

$$f: n \mapsto f(n) : \mathbb{N} \to Y$$

si scrive più semplicemente $\{f_n\}$ o, più estesamente, visualizzandone l'andamento: $f_0, f_1, f_2, \dots, f_n, \dots$.

La variabile n prende il nome di *indice*. Si impiegano poi di preferenza le lettere $a, b, \dots$ per indicare le successioni, anziché $f, g, \dots$ tradizionalmente riservate alle funzioni in generale.

Esempi

4.10. Sia $\{a_n\}$ definita da

$$n \mapsto a_n = \frac{1}{n+1} \ , \ n \in \mathbb{N} .$$

Elencando gli elementi dell'immagine avremo:

$$1, \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{n}, \frac{1}{n+1}, \dots .$$

Il grafico di $\{a_n\}$ è un sottoinsieme di $\mathbb{N} \times \mathbb{R}$. Alcuni punti sono indicati nella figura 1.10.

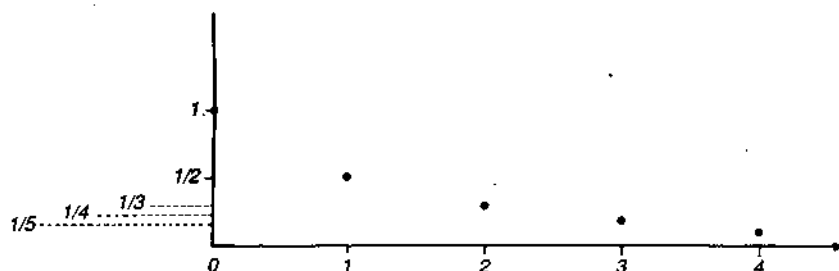



Fig. 1.10

4.11. Sia $\{a_n\}$ definita da

$$n \mapsto a_n = (-1)^n \ , \ n \in \mathbb{N} \ .$$

Avremo la sequenza

$$+1, -1, +1, -1, \ldots, (-1)^n, \ldots$$

ovvero $+1$ quando n è pari e -1 quando n è dispari.

4.12. Alcune successioni sono definite solo da un certo intero n_0 in poi. Una notazione più precisa è allora $\{a_n\}_{n>n_0}$.
Sia ad esempio $\{a_n\}_{n\geq 2}$ definita da

$$n \mapsto \sqrt{n^2 - 4} \ .$$

L'andamento di a_n è il seguente:

$$0, \sqrt{5}, \ \sqrt{12}, \ \sqrt{21}, \ldots, \sqrt{n^2 - 4}, \ldots \ .$$

Multifunzioni - Siano A e B due insiemi e $\mathcal{P}(B)$ l'insieme delle parti di B. Una funzione F da A in $\mathcal{P}(B)$, cioè:

$$F: A \to \mathcal{P}(B) \ ,$$

associa ad ogni elemento di A un sottoinsieme di B.
Una corrispondenza

elemento di $A \mapsto$ sottoinsieme di B

è chiamata *multifunzione* da A in B (o anche funzione a più valori, funzione multivoca).

Esempi

4.13. Siano $A = B = \mathbb{R}$.

Definiamo per ogni numero reale x,

$$F(x) = \{x, -x\} .$$

Cioè, ad ogni elemento $x \in \mathbb{R}$ è associato il sottoinsieme di $\mathbb{R}$ dato dai due punti x e $-x$.

Il grafico di F è illustrato nella figura 1.11.

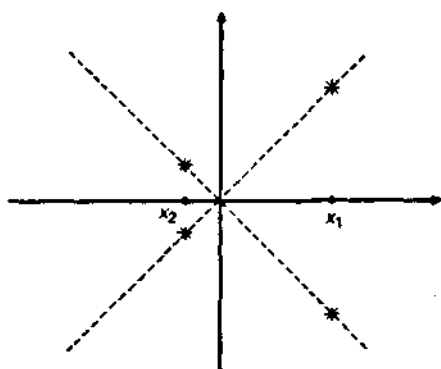



Fig. 1.11

4.14. Siano $A = B = \mathbb{R}$.
Definiamo per ogni numero reale x,

$$F(x) = \{y \in \mathbb{R}: \sin y = x\} .$$

F è una multifunzione da A in B. Osserviamo che $F(x)$ è costituito dai punti indicati in figura 1.12.

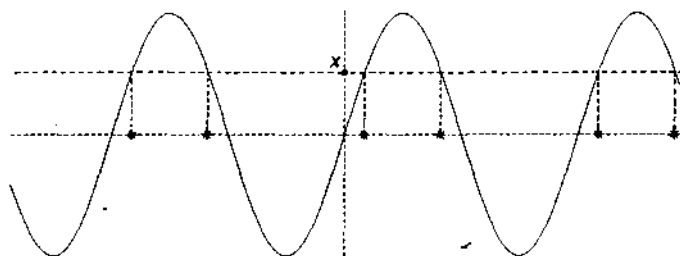



Fig. 1.12

4.3 Funzione composta

Definiamo ora un'operazione fondamentale sulle funzioni: il *prodotto di composizione* (o semplicemente *composizione*).

Definizione 4.2 - *Siano date $f: X \to Z$ e $g: W \to Y$; se risulta*

$$\operatorname{im}(f) \subseteq \operatorname{dom}(g) ,$$

ad ogni $x \in X$ f associa un unico elemento $f(x) \in \operatorname{im}(f)$ e poiché questo $f(x)$ è un elemento del $\operatorname{dom}(g)$, ad esso la funzione g associa un unico elemento $y \in Y$ dato da: $g(f(x))$. Si dice allora che g è componibile con f e risulta così definita una funzione:

$$h: x \mapsto g(f(x)) : X \to Y$$

che si dice composta di g e f e si indica con $g \circ f$ (si legga: "*g composto f*").

***Esempio* 4.15** - Siano

$$f: x \mapsto \sin x: \mathbb{R} \to \mathbb{R} \quad \text{e} \quad g: x \to \sqrt[3]{x}: \mathbb{R} \to \mathbb{R} .$$

Allora abbiamo:

$$g \circ f: x \mapsto \sqrt[3]{\sin x}: \mathbb{R} \to \mathbb{R} .$$

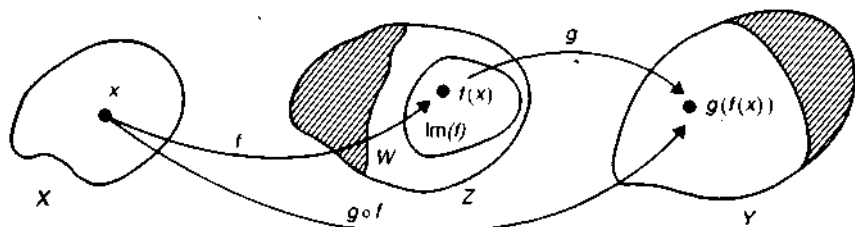



Fig. 1.13

Se g è componibile con f non è detto evidentemente che f sia componibile con g ma anche quando ciò avviene, risulterà in generale $g \circ f \neq f \circ g$, cioè questo prodotto non è commutativo.

***Esempio* 4.16** - Se f e g sono le funzioni dell'esempio precedente, risulta $f \circ g$: $x \mapsto \sin \sqrt[3]{x}$.

Se $X = Y$ la funzione $f: X \to X$ è certamente componibile con sé stessa e scriveremo f^2 invece di $f \circ f$. Così, se $f: x \mapsto \sin x$, sarà $f^2: x \mapsto \sin(\sin x)$ che dovrebbe potersi scrivere anche: $\sin^2 x$; ma quest'ultima notazione purtroppo viene spesso usata (scorrettamente) per indicare $(\sin x)^2$. Si faccia attenzione a non confondere la funzione $f^2(x)$ con la funzione $(f(x))^2$!

Il prodotto di composizione si può ovviamente estendere al caso di più di due fattori: $f \circ g \circ h \ldots$ etc.

L'allievo provi che questo prodotto è *associativo*.

4.4 Funzioni iniettive e suriettive. Funzione inversa

Definizione 4.3 - *Sia $f: X \to Y$; se avviene che* $\mathrm{im}(f) = Y$ *la funzione si dice* **suriettiva** (si dice anche che è una funzione da X *su* Y invece che *in* Y).

Diremo invece che f è **iniettiva** *se avviene che:*

$$\forall x_1 \in X \ , \ \forall x_2 \in X : f(x_1) = f(x_2) \Longrightarrow x_1 = x_2 \ ;$$

cioè *due elementi distinti x_1 e x_2 del dominio non possono avere la stessa immagine:* tali funzioni sono dette anche $1 \leftrightarrow 1$ (si legge: uno a uno). Infine, *una funzione che sia insieme iniettiva e suriettiva si dirà* **biiettiva**.

Naturalmente, qualunque funzione potrebbe rendersi suriettiva scegliendo il codominio coincidente con l'immagine; ma ciò non è semplice da farsi, poiché, come abbiamo già osservato, può essere complicato determinare l'immagine.

La proprietà di iniettività è, come subito vedremo, di grande importanza. Talvolta una funzione, che non è iniettiva sul dominio assegnato, può diventarlo su un dominio più piccolo, cioè passando a una sua restrizione.

***Esempio* 4.17** - La $f: x \mapsto \sin x : \mathbb{R} \to \mathbb{R}$ non è né suriettiva né iniettiva; la stessa corrispondenza $x \mapsto \sin x$ considerata da $\mathbb{R}$ in $\{y \in \mathbb{R}: -1 \leq y \leq 1\}$ diventa suriettiva; la sua restrizione all'insieme $\{x \in \mathbb{R}: -\pi/2 \leq x \leq \pi/2\}$ è iniettiva.

Il grafico di una funzione $f: X \to Y$, con X e Y sottoinsiemi di $\mathbb{R}$, che sia *iniettiva* ha una semplice proprietà geometrica: il grafico di f interseca ogni retta parallela all'asse delle ascisse al massimo in un punto.

Suriettività ed iniettività sono strettamente collegate con il problema di risolvere equazioni. Riferendosi, ad esempio, ad una equazione del tipo $f(x) = y$ con $f: \mathbb{R} \to \mathbb{R}$, dove x è l'incognita e y il dato, la suriettività di f significa che l'equazione ha *almeno una soluzione per ogni y*, mentre l'iniettività di f significa che *la soluzione, se c'è, è unica*.

Funzione inversa.

Sia data $f: X \to Y$. Preso $y \in f(X)$, ad esso risultano associati uno o più elementi $x \in X$ tali che $f(x) = y$ (l'insieme di questi elementi x viene detto la **controimmagine** di y). Pertanto questa relazione, che associa ad y la sua controimmagine, non è in generale una funzione da $f(X)$ in X (*); *lo è se e solo se f è iniettiva*, poiché allora la controimmagine di un elemento $y \in f(X)$ è costituita da un solo elemento $x \in X$. In tal caso *risulta definita una funzione, il cui dominio* è $f(X)$ *e la cui immagine è X: essa associa ad un dato $y \in f(X)$ quell'unico elemento $x \in X$ tale che $f(x) = y$*. Tale funzione si chiama **inversa** di f e si indica con f^{-1}.

(*) È però una funzione da $f(X)$ in $\mathcal{P}(X)$ ovvero, secondo la terminologia del paragrafo 4.2, una multifunzione da $f(X)$ in X.

Attenzione a non confondere la funzione $f^{-1}(x)$ con la funzione $f(x)^{-1} = 1/f(x)$! Per esempio, se f è l'identità in X, $f = I_X$, la sua inversa è essa stessa: $f^{-1} = f$.

Dunque *una funzione $f: X \to Y$ è invertibile* (cioè esiste la sua inversa) *se e solo se f è iniettiva, ovvero se f, considerata da X su $f(X)$, è biiettiva.*

Dalla definizione data segue immediatamente che, se f è invertibile, risulta componibile con la sua inversa sia a destra che a sinistra e si ha:

$$f \circ f^{-1} = I_{f(X)} \qquad f^{-1} \circ f = I_X$$

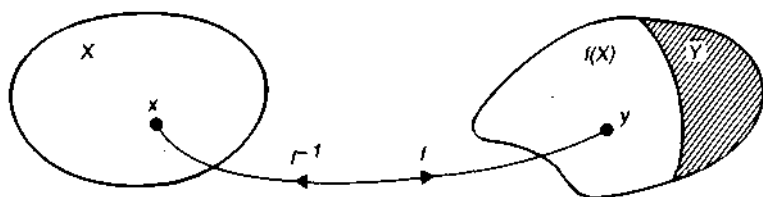



Fig. 1.14

Lo studente può anche dimostrare, per esercizio, che l'inversa di una funzione data f è pure invertibile e

$$(f^{-1})^{-1} = f\ .$$

Se dunque f è invertibile, la corrispondenza da essa determinata tra X e $f(X)$ è *biunivoca*; si dice anche che gli insiemi X e $f(X)$ sono *identificati* attraverso la f.

Esempi

4.18. La funzione, già considerata nell'esempio 4.17. definita da:

$$f: x \mapsto \sin x: A \to B$$

dove

$$A = \left\{x \in \mathbb{R}\ :\ -\frac{\pi}{2} \leq x \leq \frac{\pi}{2}\right\} \quad \text{e} \quad B = \{x \in \mathbb{R}\ :\ -1 \leq x \leq 1\}$$

è suriettiva e iniettiva. La sua funzione inversa $f^{-1}: B \to A$ si chiama *arcoseno* e si indica con il simbolo *arcsin*.

4.19. La funzione

$$f: x \mapsto x^3: \mathbb{R} \to \mathbb{R}$$

è suriettiva e iniettiva. La sua funzione inversa, $f^{-1}: \mathbb{R} \to \mathbb{R}$, si chiama *radice cubica* e si indica con il simbolo $\sqrt[3]{\ }$.

Esercizi

1. Sia P l'insieme dei parallelepipedi ed f la funzione da P in $\mathbb{R}$ data assegnando ad ogni parallelepipedo il suo volume. Esaminare iniettività e suriettività di tale funzione. Eventualmente modificando dominio e codominio di f si può rendere l'applicazione biiettiva?

2.* *Proiezione stereografica della sfera.*
Indichiamo con S^2 la superficie sferica unitaria, sottoinsieme di $\mathbb{R}^3$. Cioè

$$S^2 := \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1^2 + x_2^2 + x_3^2 = 1\} .$$

Sia $N = (0,0,1)$ il polo Nord. Definiamo la funzione $f: \mathbb{R}^2 \to S^2$ associando ad ogni punto $y = (y_1, y_2) \in \mathbb{R}^2$ il punto $f(y) = (x_1, x_2, x_3) \in S^2$, intersezione di S^2 con la retta congiungente y con N, diverso da N.

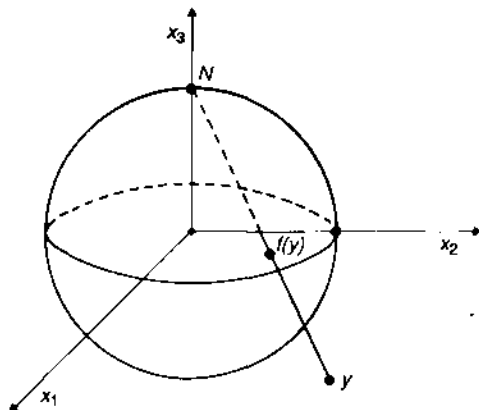



Fig. 1.15

Studiare iniettività e suriettività di f.
Verificare che l'espressione analitica di f è la seguente:

$$x_1 = \frac{2y_1}{1 + y_1^2 + y_2^2} \ , \ x_2 = \frac{2y_2}{1 + y_1^2 + y_2^2} \ , \ x_3 = 1 - \frac{2}{1 + y_1^2 + y_2^2} \ .$$

Trovare l'espressione analitica della funzione inversa f^{-1}.

3. Verificare che $f: x \mapsto \cos x : A \to B$
dove

$$A = \{x \in \mathbb{R} : 0 \le x \le \pi\}$$

e

$$B = \{x \in \mathbb{R}, -1 \le x \le 1\}$$

è suriettiva ed iniettiva. La sua funzione inversa si chiama *arcocoseno* e si indica con il simbolo *arccos*.

4. Verificare che $f: x \mapsto \operatorname{tg} x : A \to B$

dove

$$A = \left\{ x \in \mathbb{R}; -\frac{\pi}{2} < x < \frac{\pi}{2} \right\}$$

e

$$B = \mathbb{R}$$

è suriettiva ed iniettiva. La sua funzione inversa si chiama *arcotangente* e si indica con il simbolo *arctg*.

5. Siano $f(x) = 1/(1 + x^4)$ e $g(x) = x^2$.
Determinare $f \circ g$ e $g \circ f$.

6. Sia $f: \mathbb{R} \to \mathbb{R}$ definita da

$$f(x) = \begin{cases} 1 & \text{se } x \geq 0 \\ x^2 & \text{se } x < 0 \, . \end{cases}$$

Determinare $f \circ f$.

7. Sia $f: \mathbb{R}_+ \to \mathbb{R}_+$ definita da $f(x) = 1/x$.
Verificare che f è biiettiva e che $f^{-1} = f$.

8. Se p è un enunciato, poniamo $p = 0$ se p è falso e $p = 1$ se p è vero.
La funzione $f: \{0,1\} \to \{0,1\}$ definita da $f(p) = 1 - p$ "simula" la negazione "$\sim p$". Infatti $f(0) = 1$ e $f(1) = 0$.
Consideriamo ora le seguenti funzioni $f: \{0,1\} \times \{0,1\} \to \{0,1\}$;

a) $f(p,q) = p \cdot q$

b) $f(p,q) = p + q - p \cdot q$

c) $f(p,q) = 1 - p + p \cdot q$.

Quali enunciati simulano?
Se $f(p,q) = 1$ per ogni $(p,q) \in \{0,1\} \times \{0,1\}$ a quale tipo di enunciato corrisponde f?

9. Sia a_n la successione definita da:

$$a_n = n \bmod 4 \ (= \text{resto della divisione di } n \text{ per } 4) \ .$$

Descrivere l'andamento della successione a_n.
Qual è l'ultima cifra di $3^{(10^{200})}$?

5. INSIEMI FINITI

5.1 Numeri cardinali. Numeri naturali

La domanda: "quanti sono gli elementi di un insieme?", conduce al concetto di *cardinalità* di un insieme.

Soffermiamoci un momento sulla nozione di corrispondenza biunivoca tra insiemi, appena definita. In accordo con quanto detto sopra, due insiemi non vuoti, A e B, sono in *corrispondenza biunivoca* se esiste una funzione biiettiva da A su B.

Definizione 5.1 - *Si dice che due insiemi non vuoti A e B sono* **equipotenti** *se sono in corrispondenza biunivoca. Scriveremo $A \leftrightarrow B$.*

La proprietà di equipotenza è riflessiva, simmetrica e transitiva come è facile verificare. Si sarebbe perciò tentati di affermare che essa è una "relazione di equivalenza nell'insieme di tutti gli insiemi", concetto che però è contraddittorio (si veda la discussione alla fine della sezione 2).

Pur non potendo considerare l'equipotenza come una relazione di equivalenza introduciamo, dato un insieme non vuoto A, la classe di tutti gli insiemi equipotenti ad A. Questa classe "identifica" l'unica proprietà comune a tutti i suoi elementi: quella di essere in corrispondenza biunivoca tra loro. Ad essa Cantor diede il nome di **numero cardinale** di A (o anche **potenza** di A).

In altre parole:

Definizione 5.2 - *Si chiama cardinalità (o potenza o numero cardinale) di un insieme non vuoto A, e si indica con* $\mathrm{card}(A)$ *la classe degli insiemi equipotenti ad A.*

Possiamo assumere che l'insieme vuoto $\emptyset$ definisca una classe di equipotenza $\mathrm{card}(\emptyset)$.

Per mezzo del concetto di cardinalità possiamo arrivare ad una definizione di numero naturale nel modo che segue.

Per ogni insieme A definiamo il *successivo* di A, e lo indichiamo con A^+, come l'insieme ottenuto aggiungendo A stesso ai singoli elementi di A, cioè, in simboli:

$$A^+ := A \cup \{A\} = \{A, \{A\}\} .$$

A questo punto possiamo definire il numero naturale 0 ponendo:

$$0 := \mathrm{card}(\emptyset)$$

e poi 1 come la cardinalità del successivo di $\emptyset$ (in breve: 1 è il successivo di 0) cioè:

$$1 := \mathrm{card}(\emptyset^+) = \mathrm{card}(\{\emptyset\}) = \mathrm{card}(\{0\})$$

e poi 2 come il successivo di 1,

$$2 := \mathrm{card}(\{\emptyset\}^+) = \mathrm{card}(\{\emptyset, \{\emptyset\}\}) = \mathrm{card}(\{0, 1\})$$

e così via.

Per questa strada la nozione di *numero* viene ricondotta a quella di funzione e, in ultima analisi, a quella di insieme.

Naturalmente la nozione di *classe di equipotenza* è un concetto delicato; per introdurlo correttamente ed evitare contraddizioni logiche occorrerebbe far ricorso alla teoria assiomatica, che è al di là dei nostri scopi.

Tuttavia può essere istruttivo vedere come da uno degli assiomi della teoria si possa costruire l'*insieme* $\mathbb{N}$ *dei numeri naturali.*

Assioma dell'infinito: "*esiste un insieme che contiene* $\emptyset$ *e contiene il successivo di ogni suo elemento*".

Un tale insieme viene anche detto *induttivo*. Pertanto *ogni* insieme induttivo contiene lo 0 e tutti i suoi successivi. È pure semplice mostrare che l'*intersezione di due o più insiemi induttivi è un insieme induttivo.* Allora, considerato un insieme induttivo A, possiamo definire l'*insieme di tutti i numeri naturali, che indicheremo con* $\mathbb{N}$, *come l'intersezione di tutti i sottoinsiemi induttivi di* A.

La definizione, come subito si accerta, non dipende dall'insieme A scelto, poiché se B è un altro insieme induttivo, $A \cap B$ è pure induttivo e, siccome è contenuto in A, è uno dei sottoinsiemi che entrano nella definizione di $\mathbb{N}$.

Definizione 5.3 - *Un insieme* X *si dice* **finito** *se il suo numero cardinale è un numero naturale; altrimenti si dice* **infinito**.

In questa sezione e nella prossima ci occuperemo degli insiemi finiti mentre nella sezione 7 daremo alcuni cenni degli insiemi infiniti.

Operazioni sui naturali.

Dalla definizione risulta immediatamente che l'insieme $\mathbb{N}$ dei numeri naturali è ordinato con l'ordinamento indotto dalla relazione di appartenenza "$\in$".

Ad esempio $0 < 1$ (zero minore di 1) equivale a $\emptyset \in \{\emptyset\}$, così $1 < 3$ equivale a $\{\emptyset\} \in \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$ e così via.

Risulta anche che esso è *bene ordinato* (cioè ogni sottoinsieme non vuoto di $\mathbb{N}$ ha elemento minimo), dunque totalmente ordinato.

È facile verificare che se $A \subseteq B$ e $\text{card}(A) = m$, $\text{card}(B) = n$ allora $n \geq m$ (esercizio).

Introduciamo ora le operazioni fondamentali sui numeri naturali che chiameremo anche *cardinali finiti*.

Può essere interessante osservare che, con i moderni metodi didattici, nelle scuole elementari si introducono le operazioni aritmetiche seguendo le stesse idee.

Siano $m, n \in \mathbb{N}$ ed A, B due insiemi con $\text{card}(A) = m$ e $\text{card}(B) = n$.

1. *Somma:* si scelgono A e B disgiunti e si pone

$$n + m := \text{card}(A \cup B)$$

2. *Prodotto:*

$$m \cdot n := \text{card}(A \times B)$$

3. *Differenza:* si scelgono A e B in modo che $A \subseteq B$ e si pone

$$n - m := \text{card}(B \setminus A) .$$

Osserviamo che, essendo $A \subseteq B$ si ha $n \geq m$.

4. *Quoziente:* è possibile solo se, ad esempio, $m \leq n$ e se si può assegnare in B una partizione (cfr. definizione 3.4) in classi disgiunte in modo che ciascuna classe abbia cardinalità m.

Se r è la relazione di equivalenza in B determinata dalla partizione poniamo

$$\frac{n}{m} := \text{card}(B/r) .$$

Dalla definizione di somma segue che, se $n \in \mathbb{N}$, il successivo di n, n^+, è esattamente $n + 1$.

Occorre naturalmente verificare che le operazioni così definite *non dipendono* dalla coppia di insiemi A e B scelta. Tralasciamo la noiosa verifica.

Lasciamo poi allo studente verificare che le operazioni di somma e prodotto sopra definite godono della proprietà commutativa, associativa, distributiva.

Osserviamo infine che, mentre le operazioni di somma e prodotto possono definirsi senza restrizione alcuna sull'insieme dei numeri naturali, quelle di differenza e quoziente hanno senso solo in certi casi; la necessità di rimuovere queste restrizioni è la ragione principale che ci porterà ad ampliare l'insieme numerico $\mathbb{N}$ ora introdotto.

5.2 Il principio di induzione

Il concetto di numero naturale era noto, in forma intuitiva, fin dall'antichità, ma la prima presentazione assiomatica di questi numeri risale soltanto alla fine del secolo scorso (1890 circa) e si deve ad un matematico italiano, G. Peano. Egli pose alla base della sua teoria dei numeri naturali i cinque assiomi seguenti, dove "numero", "zero" e "successivo" sono da considerarsi nozioni primitive.

Lo scopo di Peano era quello di evitare il ricorso alle "classi di equipotenza" introdotte da Cantor, per il motivo indicato nel paragrafo precedente.

Se tuttavia si vuole rimanere all'interno della teoria degli insiemi occorre costruire un insieme che verifichi gli assiomi oppure *postularne* l'esistenza. In ciascuno dei casi si tratta praticamente di far ricorso alla teoria assiomatica degli insiemi.

Gli assiomi di Peano sono i seguenti:

P_1: *zero* (in simboli: 0) *è un numero naturale*

P_2: *per ogni numero naturale n esiste un numero naturale univocamente determinato detto il successivo di n* (in simboli n^+)

P_3: *0 non è il successivo di alcun numero naturale*

P_4: *numeri naturali diversi hanno successivi diversi*

P_5: **Principio di induzione:** *sia $P(n)$ un predicato riguardante il numero naturale n; se:*

i) *$P(0)$ è vera*

ii) *per ogni intero naturale n: $P(0), P(1), \ldots, P(n)$ vere $\Longrightarrow P(n^+)$ vera;*

allora l'affermazione $P(n)$ è vera per ogni numero naturale.

Dagli assiomi si possono poi definire l'ordinamento, le operazioni elementari e verificare la validità delle solite regole di calcolo. In particolare si ricava ancora $n^+ = n + 1$.

Ritorniamo ora all'insieme $\mathbb{N}$ dei numeri naturali, così come definito nel paragrafo 5.1.

È immediato verificare che $\mathbb{N}$ soddisfa gli assiomi P_1, P_2, P_3, P_4.

Verifichiamo ora che soddisfa anche l'assioma P_5, che costituisce uno degli strumenti fondamentali della matematica.

Cominciamo con una formulazione equivalente del principio.

Se $P(n)$ è un predicato definito per $n \in \mathbb{N}$, è definito il sottoinsieme di $\mathbb{N}$ per i cui elementi $P(n)$ è vera; in simboli

$$A := \{n \in \mathbb{N}: P(n)\} .$$

Viceversa, dato un sottoinsieme A di $\mathbb{N}$ risulta ad esso associato il predicato $P(n)$ definito da:

$$P(n): \text{“}n \in A\text{”} .$$

Per verificare che $\mathbb{N}$ soddisfa l'assioma P_5 dobbiamo dunque dimostrare il seguente teorema.

■ **Teorema 5.1** - *Sia A un sottoinsieme di $\mathbb{N}$ tale che:*

i) $0 \in A$

ii) $\forall n: n \in A \Longrightarrow n + 1 \in A$.

Allora $A = \mathbb{N}$.

Dimostrazione - Basterà mostrare che: CA (il complementare di A rispetto a $\mathbb{N}$) è vuoto. Per assurdo: CA non essendo vuoto, ed essendo bene ordinato, possiede un elemento minimo m che non è 0, poiché $0 \in A$.

Allora m è il successivo di un n, il quale deve appartenere ad A essendo m il minimo di CA. Ciò contraddice l'ipotesi ii). □

In conclusione, $\mathbb{N}$ soddisfa tutti gli assiomi di Peano ovvero, come si suol dire in logica, *ne costituisce un modello.*

Osservazione 5.1 - Se $n_0 \in \mathbb{N}$ e risulta

i) $P(n_0)$ è vera

ii) $\forall n \geq n_0: P(n) \Longrightarrow P(n + 1)$.

Allora $P(n)$ è vera per ogni $n \geq n_0$.

Il principio di induzione verrà ripetutamente usato come strumento per dimostrare teoremi; è inoltre il fondamento per le definizioni cosiddette per *ricorrenza* o *ricorsive*.

Cominciamo con un'applicazione del principio di induzione alla dimostrazione di un'importante formula.
Sottolineiamo che la ii) consiste nel dimostrare l'implicazione

$$P(n) \Longrightarrow P(n+1) \qquad \forall n \geq n_0 .$$

Occorre dunque assumere $P(n)$ vera (*ipotesi di ricorrenza*) e mostrare che, di conseguenza, anche $P(n+1)$ è vera.

***Esempio* 5.1** - (Formula per la somma di termini in progressione geometrica). Sia

$$P(n)\colon \sum_{k=0}^{n} q^k = \frac{1-q^{n+1}}{1-q} \quad \text{($q \neq 1, n \geq 0$; q si chiama \textit{ragione} della progressione).} \tag{5.1}$$

Vogliamo dimostrare che $P(n)$ è vera per ogni $n \geq 0$. Usiamo il principio di induzione, verificando i) e ii).

i) $P(0)$ si riduce all'uguaglianza $1 = 1$; quindi $P(0)$ è vera.

ii) Assumiamo che per n intero ≥ 0, $P(n)$ sia vera (ipotesi di ricorrenza). Si ha:

$$\sum_{k=0}^{n+1} q^k = \sum_{k=0}^{n} q^k + q^{n+1} \underset{\substack{\uparrow \\ \text{per l'ipotesi} \\ \text{di ricorrenza}}}{=} \frac{1-q^{n+1}}{1-q} + q^{n+1} = \frac{1-q^{n+2}}{1-q} .$$

Otteniamo dunque la formula

$$\sum_{k=0}^{n+1} q^k = \frac{1-q^{n+2}}{1-q}$$

che coincide con $P(n+1)$. Dunque $P(n) \Longrightarrow P(n+1)$, $\forall n \geq 0$.
Poiché i) e ii) sono verificate possiamo concludere che $P(n)$ è vera $\forall n \in \mathbb{N}$.

Veniamo ora al ruolo del principio di induzione nelle *definizioni ricorsive*, in particolare delle successioni.

Una successione

$$s\colon \mathbb{N} \to Y$$

si dice *definita ricorsivamente* o *per ricorrenza* se:

1) è assegnato il valore in 0; cioè $s_0 = \overline{y}$, $\overline{y} \in Y$;

2) è assegnata una funzione $f\colon \mathbb{N} \times Y \to Y$ mediante la quale s_{n+1} si ottiene da s_n; cioè

$$s_{n+1} := f(n, s_n) \; .$$

L'allievo verificherà, usando il principio di induzione, che *il procedimento* 1), 2) *illustrato sopra definisce univocamente una successione da* $\mathbb{N}$ *in* Y.

Esempi

5.2. Poniamo:

$$s_0 = 1 \; , \; s_{n+1} = \sqrt{1 + s_n} \quad (n \geq 0) \; .$$

Otteniamo la successione i cui primi termini sono:

$$1, \; \sqrt{2}, \; \sqrt{1+\sqrt{2}}, \; \sqrt{1+\sqrt{1+\sqrt{2}}}, \ldots$$

5.3. Poniamo:

$$s_0 = 1 \; , \; s_{n+1} = (n+1) \cdot s_n \quad (n \geq 0) \; .$$

Otteniamo la successione

$$1, 1, 2, 3 \cdot 2, 4 \cdot 3 \cdot 2, \ldots, n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot 1, \ldots$$

Si ottiene dunque una formula esplicita per s_n:

$$s_n = n \cdot (n-1) \cdot (n-2) \cdots 3 \cdot 2 \cdot 1 \; . \tag{5.2}$$

L'espressione a destra della (5.2) si chiama *fattoriale* di n e si indica con $n!$. Si pone, per definizione, $0! := 1$.

La successione $n!$ cresce molto rapidamente al crescere di n. Ecco i primi valori

n	0	1	2	3	4	5	6	7	8	9	10	...
$n!$	1	1	2	6	24	120	720	5'040	40'320	362'880	3'628'800	...

Più in generale si può definire una successione per *ricorrenza* col seguente procedimento:

1) si assegna il valore in $0, 1, 2, \ldots, k$ (k fissato);
2) si assegna la legge mediante la quale da $s_n, s_{n-1}, \ldots, s_{n-k}$ si ottiene s_{n+1} .

***Esempio* 5.4** - Poniamo:

$$s_0 = 0 \; , \; s_1 = 1 \quad \text{e} \quad s_{n+1} = s_n + s_{n-1} \quad (n \geq 1) \; .$$

Questa successione, i cui primi termini sono

$$0, 1, 1, 2, 3, 5, 8, 13, \ldots$$

si chiama successione di Fibonacci.

Esercizi

1. Usando il principio di induzione, dimostrare che

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2} .$$

2. Dimostrare per induzione che

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6} .$$

3. Dimostrare per induzione che

$$\sum_{k=1}^{n} (2k-1) = n^2 .$$

4.* Usando il principio di induzione dimostrare che, dati n numeri positivi $x_1, \ldots, x_n$, $n \geq 2$, tali che $x_1 \cdot x_2 \cdot \ldots \cdot x_n = 1$ si ha $x_1 + x_2 + \cdots + x_n \geq n$.
Inoltre $x_1 + x_2 + \cdots + x_n = n$ se e solo se $x_1 = x_2 = \cdots = x_n = 1$.

5. (Duale del precedente). Usando il principio di induzione dimostrare che, dati n numeri positivi $x_1, x_2, \ldots, x_n$, $n \geq 2$, tali che $x_1 + x_2 + \cdots + x_n = n$ allora si ha $x_1 \cdot x_2 \cdot \ldots \cdot x_n \leq 1$.
Inoltre $x_1 \cdot x_2 \ldots x_n = 1$ se e solo se $x_1 = x_2 = \cdots = x_n = 1$.

6. Dimostrare per induzione che, dati due interi naturali m, n con $m \geq n \geq 1$, esiste un'unica coppia di naturali, q ed r, tali che

$$m = nq + r \quad \text{e} \quad 0 \leq r < n .$$

Come si chiamano q ed r?

7. Siano date n rette di un piano in posizione generica, $n \geq 2$ (2 rette qualsiasi si intersecano ma 3 qualsiasi *non* hanno un punto in comune). Sia $p(n)$ il numero di parti in cui viene suddiviso il piano. Dimostrare che

i) $p(n+1) = p(n) + (n+1)$,

ii) $p(n) = (n^2 + n + 2)/2$.

(vedi fig. 1.16).

3 rette dividono il piano in 7 parti

4 rette dividono il piano in 11 parti

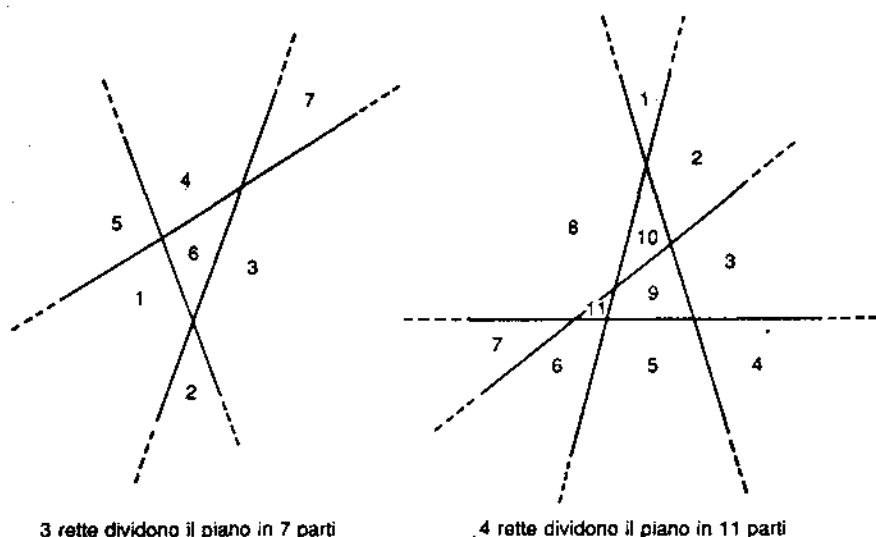



Fig. 1.16

8.* Definiamo "mappa rettilinea" un insieme finito di segmenti nel piano che colleghino n punti $P_1, P_2, \ldots, P_n$ $(n \geq 2)$ e che non si intersechino in altri punti. Sia inoltre possibile partire da uno qualunque dei punti $P_1, \ldots, P_n$ e raggiungere uno qualsiasi degli altri, muovendosi lungo segmenti della mappa (cioè la mappa è *connessa* per segmenti). Sia r il numero delle regioni in cui il piano viene suddiviso dalla mappa ed s il numero dei segmenti.

Ad esempio, nella mappa in figura 1.17 vi sono 10 punti (vertici), 11 segmenti e tre regioni (I, II, III).

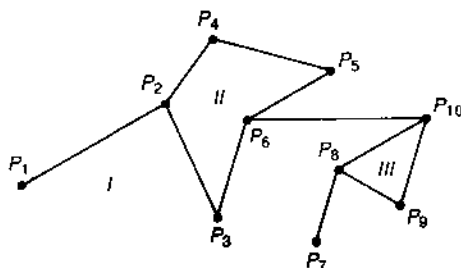



Fig. 1.17

Dimostrare per induzione sul numero s dei segmenti che

$$r + n = s + 2 \quad \text{(formula di Eulero)} .$$

[Suggerimento: verificare la formula per $s = 1$; distinguere poi il caso in cui vi è un vertice estremo di un solo arco (come P_1 o P_7 in figura) dal caso in cui da ogni vertice "partono" almeno 2 segmenti].

6. ELEMENTI DI CALCOLO COMBINATORIO

Introdurremo ora alcuni argomenti di Calcolo Combinatorio. Essi si riveleranno indispensabili durante tutto il corso di Analisi (e anche in altri corsi, come Geometria, Algebra lineare, Probabilità e Statistica, ...). Anche se le considerazioni che ora svolgeremo potrebbero trovare posto in altra parte del libro, ci sembra che sia a questo punto la loro collocazione più naturale. Fra l'altro, esse ci permetteranno di rispondere a domande che si pongono in maniera spontanea per gli *insiemi finiti*, quali, ad esempio: quanti sono gli ordinamenti che si possono introdurre in un insieme finito? Quanti sono i sottoinsiemi di cardinalità assegnata che sono inclusi in un insieme finito? Quanti sono gli elementi dell'insieme delle parti di un insieme finito? Quante sono le applicazioni di un insieme finito in un altro?...

Più precisamente, svilupperemo la teoria cercando di rispondere ai seguenti quesiti:

1) In quanti modi si possono disporre 7 persone in fila indiana?

2) Quanti sono i possibili anagrammi della parola ANAGRAMMA?

3) In quanti modi si possono scegliere 5 carte da un mazzo di 32?

4) Quante sono le quaterne di interi naturali (x_1, x_2, x_3, x_4) soluzioni della seguente equazione?

$$x_1 + x_2 + x_3 + x_4 = 7 .$$

5) Quanti sono i possibili sottoinsiemi di un insieme di n elementi?

La risposta a queste domande coinvolge gli aspetti più elementari del calcolo combinatorio, svolti nel paragrafo 6.1.

Per rispondere a quelle che seguiranno ricorreremo ad un principio più profondo, il **principio di inclusione ed esclusione**, altrimenti noto come **teorema del crivello**, presentato nel paragrafo 6.2. Infine, nel paragrafo 6.3, accenneremo ad una notevole applicazione di molti dei concetti sopra esposti.

6) Problema "*degli incontri*" o "*delle associazioni proibite*".

Ad una festa da ballo prendono parte n coppie di coniugi. In quanti modi ogni marito (moglie) può accoppiarsi in modo da *evitare* di ballare con la propria moglie (col proprio marito)?

7) In una ditta, ogni giorno, vengono assegnate n mansioni diverse ad n impiegati. In quanti modi si possono assegnare le mansioni in modo tale che *nessun* impiegato abbia la *stessa* mansione del giorno prima?

8) Data una scacchiera, in quanti modi è possibile disporre 8 torri in modo che *nessuna* di esse sia sulla diagonale bianca e che *nessuna* possa catturarne un'altra?

Una possibile configurazione è quella indicata a pagina seguente.

	T						
T							
			T				
				T			
						T	
		T					
							T
					T		

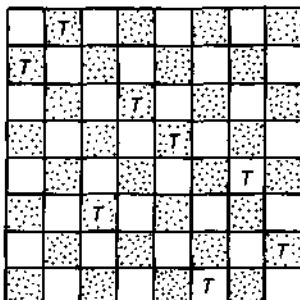



Fig. 1.18

Come vedremo, i quesiti 6, 7, 8 hanno una matrice comune.

6.1 Permutazioni, combinazioni, disposizioni

Definizione 6.1 - *Ogni ordinamento di un insieme X di n oggetti distinti prende il nome di* **permutazione** *o* **sostituzione.**

Ogni ordinamento degli elementi di X induce una *corrispondenza biunivoca* tra X e l'insieme campione $E_n := \{1, 2, \ldots, n\}$.
Per esempio, se $X = \{a, b, c\}$, ci sono 6 possibili ordinamenti:

$$\begin{array}{cccccc} abc & \begin{matrix} acb \\ \updownarrow\updownarrow\updownarrow \\ \mathbf{123} \end{matrix} & bac & bca & cab & cba\,. \end{array}$$

Viceversa, ogni corrispondenza $1 \leftrightarrow 1$ tra X ed E_n individua un particolare ordinamento degli elementi di X: tale corrispondenza infatti, etichetta gli elementi di X con indici da 1 ad n.

Problema - Quante sono le permutazioni di n oggetti distinti? ovvero: dato X con $\text{card}(X) = n$, quante sono le applicazioni biiettive di X su E_n?
Indichiamo con P_n tale numero.

■ **Teorema 6.1** - $P_n = n!$.

Dimostrazione - Consideriamo n scatole, come indicato in figura.

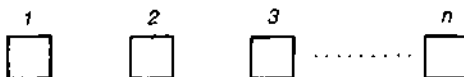



Fig. 1.19

Il numero delle permutazioni di n oggetti è uguale al numero dei modi in cui si possono collocare gli oggetti nelle scatole.
La 1^a scatola si può riempire in n modi.
Riempita la 1^a scatola si può riempire la 2^a in $(n - 1)$ modi con i restanti oggetti.

Così, il complesso delle prime 2 scatole si può riempire in $n(n-1)$ modi.
Il complesso delle prime 3 scatole si può riempire in $n(n-1)(n-2)$ modi.
Giunti alla k–esima scatola abbiamo collocato $k-1$ oggetti, per cui questa scatola si può riempire in $n-(k-1)=n-k+1$ modi; di conseguenza il complesso delle prime k scatole si può riempire in $n(n-1)(n-2)\cdots(n-k+1)$ modi.
Quando $k=n$ si ottiene $n!$ che è la tesi. □

Possiamo rispondere alla domanda 1) posta all'inizio del paragrafo 6.

Risposta alla domanda 1): il numero dei modi di disporre 7 persone in fila indiana è 7!.

Permutazioni con ripetizione. Consideriamo 7 oggetti $aaabbcc$ di cui tre uguali tra loro, due uguali fra loro e distinti dai precedenti, altri due uguali fra loro e distinti dai precedenti. È evidente che, non distinguendo più gli a fra loro, etc., il numero delle permutazioni sui 7 oggetti sarà minore di 7!. Per calcolarlo, distinguiamo con un indice gli oggetti uguali: $a_1a_2a_3b_1b_2c_1c_2$; consideriamo una qualsiasi delle 7! permutazioni di questi oggetti distinti, per esempio $b_2c_1a_3c_2a_2b_1a_1$.
Se ora togliamo l'indice delle 3 lettere a, esse possono essere permutate in uno qualunque dei 3! modi nella terna dei posti da esse occupati senza che le permutazioni si distinguano fra loro; ciò significa che le permutazioni con gli elementi $a_1a_2a_3$ non distinti tra loro risultano ora 7!/3!.
Con analogo ragionamento, se ora togliamo gli indici anche alle lettere b e c, concludiamo che le permutazioni sono 7!/3!2!2!.

Le considerazioni sopra esposte hanno carattere generale.
Sia X un insieme di n oggetti di cui

k_1 uguali tra loro

k_2 uguali tra loro e distinti dai precedenti

⋮

k_h uguali tra loro e distinti dai precedenti

con

$$k_1+k_2+\cdots+k_h=n\ .$$

Quante sono le permutazioni distinte di questi n oggetti?
Indichiamo con $P^*_{k_1,\ldots,k_h}$ il loro numero.

■ **Teorema 6.2** - $P^*_{k_1,k_2,\ldots,k_h}=\dfrac{n!}{k_1!k_2!\ldots k_h!}$

Rispondiamo ora ai quesiti 2) e 4).

Risposta al quesito 2). Gli anagrammi di "ANAGRAMMA" sono 9!/4!2! = 7560.

Risposta al quesito 4). Cerchiamo il numero di quaterne (x_1, x_2, x_3, x_4) di interi naturali soluzioni dell'equazione

$$x_1 + x_2 + x_3 + x_4 = 7 .$$

Sia (x_1, x_2, x_3, x_4) una soluzione. A questa associamo un allineamento di 3 cerchi e 7 asterischi nel modo seguente:

$x_1$asterischi [cerchio] $x_2$asterischi [cerchio] $x_3$asterischi [cerchio] $x_4$asterischi .

Ad esempio, alla soluzione $(1,3,2,1)$ è associato l'allineamento

$$\underbrace{*}_{x_1} \circ \underbrace{***}_{x_2} \circ \underbrace{**}_{x_3} \circ \underbrace{*}_{x_4} .$$

Viceversa, ogni allineamento di 3 cerchi e 7 asterischi individua una quaterna soluzione.

Ad esempio all'allineamento

$$\circ * * * * \circ * \circ * *$$

corrisponde la soluzione $(0,4,1,2)$.

Dunque, l'insieme delle quaterne cercate è in corrispondenza biunivoca con le permutazioni di 10 oggetti di cui 3 uguali tra loro e 7 uguali tra loro e distinti dai precedenti. Ne segue che il numero cercato è

$$P^*_{7,3} = \frac{10!}{7!3!} = 120 .$$

Il lettore può dimostrare che, in generale, il numero di n-uple di interi positivi $(x_1, x_2, \ldots, x_n)$ soluzioni dell'equazione

$$x_1 + x_2 + \cdots + x_n = k$$

è dato dalla formula:

$$P^*_{k,n-1} = \frac{(n+k-1)!}{k!(n-1)!} . \tag{6.1}$$

I numeri $n!/k_1!k_2!\ldots k_h!$ (n fisso, $k_1, k_2, \ldots, k_h$ variabili da 0 a n con il vincolo $k_1 + k_2 + \cdots + k_h = n$) si dicono *coefficienti polinomiali* di grado n. Essi intervengono, oltre che nello sviluppo della potenza naturale di un polinomio (vedi sotto), anche in importanti questioni di probabilità.

Il caso particolare $h = 2$ è di speciale importanza. Indicando allora k_1 con k e k_2 con $n-k$, i numeri $n!/k!(n-k)!$ (n fisso, k variabile da 0 a n) si chiamano *coefficienti binomiali* di grado n e vengono indicati con un simbolo speciale; si pone infatti:

$$\binom{n}{k} := \frac{n!}{k!(n-k)!} = \frac{n(n-1)\ldots(n-k+1)}{k!} .$$

Lo studente verificherà, con semplici calcoli, due importanti proprietà dei coefficienti binomiali, cioè:

$$\binom{n}{k} = \binom{n}{n-k}$$

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k} .$$

Quest'ultima permette di calcolare i numeri $\binom{n}{k}$ per mezzo del cosiddetto triangolo di Tartaglia; in cima al triangolo si pone il nunero $\binom{0}{0} = 1$ (per definizione), ai lati si pongono i numeri $\binom{n}{0} = \binom{n}{n} = 1$ per ogni $n \geq 1$ e allora, per $0 < k < n$, il numero $\binom{n}{k}$ viene scritto all'incrocio della n-esima riga e della k-esima colonna e risulta somma dei due numeri che si trovano nella riga precedente, quello sulla stessa colonna e quello sulla colonna precedente.

$$\begin{array}{llllll}
1 \\
1 & 1 \\
1 & 2 & 1 \\
1 & 3 & 3 & 1 \\
1 & 4 & 6 & 4 & 1 \\
1 & 5 & 10 & 10 & 5 & 1 \\
\dots & \dots & \dots & \dots & \dots & \dots
\end{array}$$

La ragione del nome dato ai numeri $\binom{n}{k}$ risulta chiara dalla seguente formula:

▪ **Teorema 6.3** - (Formula del binomio di Newton). *Se a e b sono numeri reali e n un intero naturale,*

$$(a+b)^n = a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \cdots + \binom{n}{n-1}ab^{n-1} + b^n = \\ = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k . \tag{6.2}$$

Dimostrazione - Si potrebbe ottenere per induzione. Una dimostrazione diretta, di carattere combinatorio, è la seguente.
Il prodotto $(a+b)\,(a+b)\dots(a+b)$ di n fattori si sviluppa in una somma di monomi di grado n del tipo $a^{n-k}b^k$ con $0 \leq k \leq n$. Quanti sono i monomi di un certo tipo (cioè con k fisso)? Tanti quante sono le permutazioni di n oggetti di cui k uguali a b e i rimanenti uguali ad a, cioè $\binom{n}{k}$. □

Con ragionamenti di tipo induttivo o di tipo combinatorio la formula precedente può generalizzarsi al caso di un polinomio. Vale infatti la seguente formula

(di Leibniz), la cui dimostrazione lasciamo per esercizio: se $a_1, a_2, \ldots, a_h$ sono numeri reali e n un intero naturale,

$$(a_1 + a_2 + \cdots + a_h)^n = \sum_{*} \frac{n!}{k_1!k_2!\ldots k_h!} a_1^{k_1} a_2^{k_2} \ldots a_h^{k_h}$$

dove la somma va estesa a tutte le h-uple di numeri interi non negativi $k_1, k_2, \ldots, k_h$ tali che $k_1 + k_2 + \cdots + k_h = n$.

Definizione 6.2 - *Sia X un insieme di n elementi distinti. Ogni sottoinsieme di X costituito da k elementi* $(0 \leq k \leq n)$ *viene detto* **combinazione semplice** *di n oggetti di classe k;* il loro numero si indica col simbolo $C_{n,k}$.

Ad esempio, i sottoinsiemi di 2 elementi che si possono formare da un insieme $X = \{a, b, c, d\}$ di 4 elementi sono i 6 seguenti:

$$\{a,b\} \quad \{a,c\} \quad \{a,d\} \quad \{b,c\} \quad \{b,d\} \quad \{c,d\} .$$

Il numero $C_{n,k}$ corrisponde ai diversi modi in cui si possono scegliere k elementi tra n (distinti) dati.

■ **Teorema 6.4** - $C_{n,k} = \binom{n}{k}$.

Dimostrazione - Consideriamo gli n elementi di X; un sottoinsieme di k elementi può essere individuato applicando un'etichetta bianca su k elementi di X e una nera sui rimanenti $n - k$; così facendo resta individuata una permutazione di n oggetti di cui k uguali tra loro (etichetta bianca) e $n - k$ uguali tra loro e distinti dai precedenti (etichetta nera). Viceversa, ogni permutazione di questo tipo individua un sottoinsieme di X di k elementi. Dunque questi sottoinsiemi sono tanti quante le permutazioni $P^*_{k,n-k}$, cioè $\binom{n}{k}$. □

Risposta alla domanda 3): da un mazzo di 32 carte si possono scegliere 5 carte in $\binom{32}{5}$ modi. Sono infatti le combinazioni di 5 carte scelte fra 32.

Risposta alla domanda 5): dato un insieme X di cardinalità n, qual è la cardinalità di $\mathcal{P}(X)$?

Con gli elementi di X possiamo formare $\binom{n}{1} = n$ sottoinsiemi di 1 elemento, $\binom{n}{2}$ sottoinsiemi di 2 elementi, e così via; aggiungendo a questi l'insieme vuoto e X stesso otteniamo un totale di $1 + \binom{n}{1} + \binom{n}{2} + \cdots \binom{n}{n-1} + 1$ insiemi: dalla formula di Newton con $a = b = 1$ otteniamo

$$\text{card}(\mathcal{P}(X)) = 1 + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n-1} + 1 = 2^n .$$

La notazione 2^X si usa per indicare il numero cardinale di $\mathcal{P}(X)$ anche quando X non è finito.

Definizione 6.3 - *Sia X un insieme di n elementi (distinti). Ogni raggruppamento di k elementi di X ammettendo anche ripetizioni di oggetti, viene detto* **combinazione con ripetizione** *di n oggetti di classe k.*
Qui, ovviamente, può essere anche $k > n$. Indichiamo il loro numero con $C^*_{n,k}$.

Ad esempio, le combinazioni con ripetizione di 2 elementi che si possono formare da un insieme $X = \{a, b, c, d\}$ di 4 elementi sono 10:

$$aa \quad ab \quad ac \quad ad \quad bb \quad bc \quad bd \quad cc \quad cd \quad dd\,.$$

■ **Teorema 6.5** - $C^*_{n,k} = \binom{n+k-1}{k}$.

Dimostrazione - Consideriamo l'equazione $x_1 + x_2 + \cdots + x_n = k$ dove $x_1, x_2, \ldots, x_n$ sono interi naturali. Sia $X = \{a_1, a_2, \ldots, a_n\}$.
Ogni n-upla soluzione $(x_1, x_2, \ldots, x_n)$ individua una combinazione con ripetizione data da:

a_1 ripetuto x_1 volte ; a_2 ripetuto x_2 volte ; ... ; a_n ripetuto x_n volte .

Viceversa, data una combinazione con ripetizione, se poniamo, per ogni $j = 1, \ldots, n$

$x_j :=$ numero di volte in cui a_j è ripetuto nella combinazione,

otteniamo una n-upla soluzione.
Ricordando la formula (6.1) si ottiene la tesi. □

Definizione 6.4 - *Si dice* **disposizione semplice** *di n oggetti distinti di classe k ($k \leq n$) ogni ordinamento di k oggetti comunque scelti fra gli n; due disposizioni possono differire o per gli oggetti contenuti oppure per l'ordine degli oggetti stessi.*
Il numero di queste disposizioni si indica con $D_{n,k}$.

Ad esempio, le disposizioni dei 4 oggetti a, b, c, d di classe due sono le seguenti (in numero di 12):

$$ab \quad ba \quad ac \quad ca \quad ad \quad da \quad bc \quad cb \quad bd \quad db \quad cd \quad dc\,.$$

■ **Teorema 6.6** - $D_{n,k} = n(n-1)(n-2)\ldots(n-k+1) = C_{n,k} \cdot P_k$.

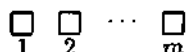
Dimostrazione - Scegliamo, fra gli n oggetti dati, k qualsiasi di essi; ciò può essere fatto in $C_{n,k}$ modi diversi. Questi k oggetti possono essere ordinati in P_k modi diversi; da cui l'asserto. □

Definizione 6.5 - *Se in una disposizione di n oggetti di classe m si possono ripetere gli oggetti* (quindi $m \gtrless n$) *parleremo di* **disposizione con ripetizione** *di n oggetti di classe m.*

Il numero di queste disposizioni viene indicato con $D^*_{n,m}$. Si trova:

■ **Teorema 6.7** - $D^*_{n,m} = n^m$.

Dimostrazione - Consideriamo m scatole:

$$\underset{1}{\square}\ \underset{2}{\square}\ \cdots\ \underset{m}{\square}$$

Possiamo riempire la 1^a scatola in n modi, la 2^a in n modi, ..., l'ultima in n modi. Dunque tutti i modi possibili di riempire le m scatole sono in numero di n^m. □

***Esempio* 6.1** - Si abbiano 13 caselle nelle quali si vogliano disporre 3 oggetti: $1, X, 2$ eventualmente ripetuti; il numero delle disposizioni è 3^{13}.

Rivediamo i teoremi 6.6 e 6.7 dal punto di vista della teoria degli insiemi finiti. Sia X un insieme di cardinalità $k \geq 1$ e Y un insieme di cardinalità $n \geq 1$. *Quante sono le applicazioni iniettive definite su X con immagine in Y?*

Siano $x_1, x_2, \ldots, x_k$ gli elementi di X. Per definire un'applicazione iniettiva f da X in Y è necessario e sufficiente: i) assegnare l'immagine di f in Y, cioè scegliere un sottoinsieme di Y costituito da k elementi (e ciò può farsi in $C_{n,k}$ modi distinti); ii) ordinare poi i k elementi di questo sottoinsieme in modo che il primo sia il corrispondente di x_1, il secondo di x_2 e così via (e ciò può farsi in P_k modi distinti). Perciò il numero delle applicazioni iniettive di X in Y è precisamente $D_{n,k}$. Si osservi che $D_{n,1} = n$ e $D_{n,n} = n!$ (in quest'ultimo caso tutte le applicazioni sono biiettive).

Concludiamo il paragrafo con un ultimo problema: sia X un insieme di cardinalità $m \geq 1$ e Y un insieme di cardinalità $n \geq 1$; quante sono le applicazioni con dominio X e codominio Y?

Siano $x_1, x_2, \ldots, x_m$ gli elementi di X; per assegnare un'applicazione f da X in Y si può procedere così: prima assegniamo il corrispondente di x_1, $f(x_1)$: ciò può essere fatto in n modi diversi; poi assegniamo il corrispondente di x_2, $f(x_2)$: ciò può essere fatto ancora in n modi diversi (poiché l'applicazione non è necessariamente iniettiva); poi assegniamo il corrispondente di x_3, ancora in n modi diversi, e così via; il numero di tutte le possibili applicazioni risulterà pertanto uguale al numero delle disposizioni con ripetizione di n elementi in m posti e cioè n^m.

La notazione Y^X si usa per indicare il numero cardinale dell'insieme di tutte le applicazioni con dominio X e codominio Y, anche quando questi insiemi non siano finiti.

*6.2 Il principio di inclusione ed esclusione

Questo principio, noto anche come *teorema del crivello*, è di così elementare comprensibilità da rendere sorprendente la sua potenza.

Nella sua forma più semplice consiste nella seguente formula. Siano A_1 e A_2

insiemi finiti ed indichiamo con $|A_1|$, $|A_2|$ le loro cardinalità.
Allora

$$|A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2| \,. \tag{6.3}$$

La (6.3) è evidente poiché $|A_1| + |A_2|$ "conta" 2 volte ciascun elemento dell'intersezione.

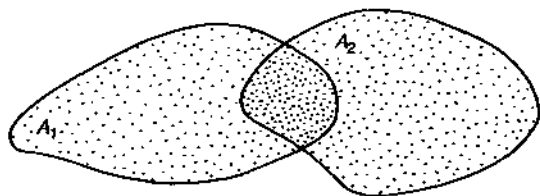



Fig. 1.20

***Esempio* 6.2** - Se

$A_1 = \{1,2,3,4,5\}$

e

$A_2 = \{3,4,5,6,7,8,9\}$,

si ha:

$$|A_1| = 5 \ , \ |A_2| = 7 \ , \ |A_1 \cap A_2| = 3 \ , \ |A_1 \cup A_2| = 9 \,.$$

Dunque

$$9 = |A_1 \cup A_2| = |A_1| + |A_2| - |A_1 \cap A_2| = 5 + 7 - 3 \,.$$

Proviamo ora con 3 insiemi A_1, A_2, A_3. Come deve essere modificata la formula (6.3)?

La formula

$$|A_1 \cup A_2 \cup A_3| = |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_2 \cap A_3| - |A_1 \cap A_3|$$

è sbagliata poiché, come è evidente dalla figura che segue, i punti dell'intersezione $A_1 \cap A_2 \cap A_3$ (contrassegnati con le crocette) vengono contati 3 volte in $|A_1| + |A_2| + |A_3|$ e poi sottratti 3 volte nei termini negativi.

Occorre dunque aggiungere i punti di $A_1 \cap A_2 \cap A_3$.

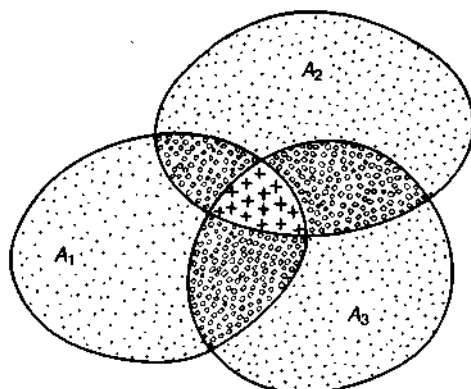



Fig. 1.21

La formula corretta è perciò la seguente:

$$|A_1 \cup A_2 \cup A_3| = |A_1| + |A_2| + |A_3| - |A_1 \cap A_2| - |A_2 \cap A_3| - \\ - |A_1 \cap A_3| + |A_1 \cap A_2 \cap A_3| \,. \tag{2}$$

Generalizziamo.
Siano $A_1, A_2, \ldots, A_n$ n insiemi.
Poniamo:

$$\nu_1 := |A_1| + |A_2| + \cdots + |A_n|$$

$$\nu_2 := |A_1 \cap A_2| + |A_1 \cap A_3| + \cdots + |A_{n-2} \cap A_n| + |A_{n-1} \cap A_n|$$

(in questa somma vi sono $\binom{n}{2}$ addendi)

$$\nu_3 := |A_1 \cap A_2 \cap A_3| + |A_1 \cap A_2 \cap A_4| + \cdots + |A_{n-2} \cap A_{n-1} \cap A_n|$$

(in questa somma vi sono $\binom{n}{3}$ addendi)

$$\vdots$$

$$\nu_n := |A_1 \cap A_2 \cap \ldots \cap A_n| .$$

■ **Teorema 6.8** - (di inclusione-esclusione o del crivello).
Sia $\nu := |A_1 \cup A_2 \cup \ldots \cup A_n|$; *vale la formula*:

$$\nu = \nu_1 - \nu_2 + \nu_3 - \cdots + (-1)^{n+1}\nu_n \,. \tag{6.4}$$

Dimostrazione - È sufficiente far vedere che ogni elemento $x \in A_1 \cup A_2 \cup \ldots \cup A_n$ è contato *esattamente una volta* nella somma a destra della (6.4).

Ora, ogni elemento $x \in A_1 \cup A_2 \cup \ldots \cup A_n$ apparterrà ad alcuni, possibilmente ad uno solo, tra gli insiemi $A_1, \ldots, A_n$. Possiamo supporre che $x \in A_1 \cap A_2 \cap \ldots \cap A_t$, $t \geq 1$, e che x non appartenga agli altri insiemi $A_{t+1}, A_{t+2}, \ldots, A_n$.

Osserviamo ora che:

nel numero ν_1 l'elemento x è contato t volte;
nel numero ν_2 l'elemento x è contato $\binom{t}{2}$ volte;

$\vdots$

nel numero ν_k $(k \leq t)$, x è contato $\binom{t}{k}$ volte;

$\vdots$

nel numero ν_t x è contato 1 volta;
infine nei numeri ν_k con $k > t$ l'elemento x *non* è contato.
Dunque nel numero a destra della (6.4) x è contato un numero di volte dato da:

$$t - \binom{t}{2} + \binom{t}{3} - \dots + (-1)^{t+1} =$$

$$= 1 - 1 + t - \binom{t}{2} + \binom{t}{3} - \dots + (-1)^{t+1} =$$

$$= 1 - \left[1 - t + \binom{t}{2} - \binom{t}{3} + \dots + (-1)^{t}\right] = \text{(formula del binomio)}$$

$$= 1 - [1 - 1]^t = 1 \ \square$$

Con il teorema 6.8 possiamo rispondere alle domande 6), 7) ed 8) poste all'inizio del paragrafo 6.1

Premettiamo un'osservazione importante.

Osservazione 6.1 - Nei casi concreti sono assegnati un insieme H ed n suoi sottoinsiemi $A_1, A_2, \dots, A_n$ tali che $A_1 \cup A_2 \cup \dots \cup A_n = A \subset H$.

Ognuno di questi sottoinsiemi è definito da una proprietà: ad esempio, A_j sia definito dalla proprietà P_j.

Allora le varie intersezioni che compaiono nella definizione dei numeri ν_j, si interpretano così:

$A_j \cap A_i$ è l'insieme dei punti che hanno le proprietà P_j e P_i;

$A_j \cap A_i \cap A_k$ è l'insieme dei punti che hanno le proprietà P_j, P_i, P_k e così via.

A si interpreta come l'insieme dei punti che hanno *almeno una tra le proprietà* $P_1, P_2, \dots, P_n$.

La formula (6.4) permette di determinare la cardinalità di A, ma *anche la cardinalità* dell'insieme dei punti di H che *non hanno alcuna* delle proprietà $P_1, P_2, \dots, P_n$; tale insieme non è altro che $H \setminus A$ e quindi, essendo $A \subset H$, si ha

$$|H \setminus A| = |H| - |A| = |H| - \nu$$

dove ν è dato dalla (6.4).

Veniamo ora al problema 6), o delle *associazioni proibite*.

Qui si hanno k coppie di coniugi. Si tratta di trovare il numero dei modi possibili di accoppiare gli n mariti ad una donna diversa dalla propria moglie.

Indichiamo i mariti con U_j, e le mogli con D_j, $j = 1, \dots, n$.

Elenchiamo le mogli secondo l'indice crescente

$$D_1 \quad D_2 \quad D_3 \quad \dots \quad D_n \; .$$

Ogni elenco di coppie (accoppiamento) si otterrà fissando una permutazione dei mariti:

$$U_{j_1}, U_{j_2}, U_{j_3}, \dots, U_{j_n} \; .$$

Dunque, tutti gli accoppiamenti possibili sono in numero di $n!$.

Secondo le notazioni dell'osservazione 6.1 sia H l'insieme di tutti i possibili accoppiamenti.

Diremo ora che un accoppiamento ha la proprietà P_j se alla donna D_j corrisponde il proprio marito U_j. Ognuna delle proprietà P_j definisce un sottoinsieme A_j di H.

Allora $A_j \cap A_i$ sarà il sottoinsieme costituito dagli accoppiamenti tali che a D_j corrisponde U_j e a D_i corrisponde U_i.

Così, $A_j \cap A_i \cap A_k$ sarà il sottoinsieme costituito dagli accoppiamenti tali che a D_j corrisponde U_j, a D_i corrisponde U_i e a D_k corrisponde U_k.

Analogamente per le altre intersezioni.

Ciò che interessa è trovare il numero degli accoppiamenti che *non* hanno *alcuna* delle proprietà P_j.

Applicando il teorema 6.8 abbiamo che tale numero è dato da

$$n! - \nu_1 + \nu_2 - \nu_3 + \cdots + (-1)^n \nu_n \; .$$

Si tratta ora di calcolare i numeri ν_j.

Si ha:

$$\nu_1 = |A_1| + |A_2| + \cdots + |A_n| \; .$$

Ogni elemento di A_j è una permutazione con il posto j-esimo bloccato; dunque $|A_j| = (n-1)!$, cosicché $\nu_1 = n(n-1)! = n!$.

Passiamo a ν_2; si ha:

$$\nu_2 = |A_1 \cap A_2| + |A_1 \cap A_3| + \cdots + |A_{n-1} \cap A_n| \; .$$

Ogni elemento di $A_j \cap A_i$ è una permutazione con i posti j ed i bloccati; dunque $|A_j \cap A_i| = (n-2)!$. Essendo la somma in ν_2 composta di $\binom{n}{2}$ termini si ottiene

$$\nu_2 = \binom{n}{2}(n-2)! \; .$$

Con gli stessi ragionamenti si ricava

$$\nu_j = \binom{n}{j}(n-j)! = \frac{n!}{j!}\,.$$

Conclusione: il numero cercato è, raccogliendo $n!$:

$$n!\left\{1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!}\right\}.$$

Se, per esempio, $n = 4$, avremo:

$$24\left\{\frac{1}{2} - \frac{1}{6} + \frac{1}{24}\right\} = 12 - 4 + 1 = 9 \text{ modi diversi.}$$

I problemi 7), 8) si risolvono esattamente nella stessa maniera ed invitiamo il lettore a convincersene, inquadrandoli nello schema generale indicato nell'osservazione 6.1.

*6.3 Probabilità in spazi finiti

Esperimenti (o prove) come il lancio di una moneta o di un dado, l'estrazione di una pallina da un'urna contenente n palline numerate o di una carta da un mazzo, etc. sono tipici *fenomeni casuali*, il cui esito cioè non è determinato dalle informazioni in possesso dello sperimentatore. Un siffatto esperimento ha perciò una molteplicità di esiti possibili, ognuno dei quali viene detto *evento elementare*: uno e uno solo di questi eventi elementari deve verificarsi necessariamente. Nel lancio di un dado, per esempio, questi eventi elementari sono i numeri da 1 a 6 e perciò formano l'insieme $\{1,2,3,4,5,6\}$; nel lancio di due dadi A e B, gli eventi elementari sono le coppie ordinate (i,j) con i e j numeri da 1 a 6: i è il numero uscito sul dado A, j quello su B; tali eventi formano un insieme costituito da 36 elementi.

Nella grande maggioranza delle applicazioni, si può ritenere che gli esiti di un esperimento casuale formino un insieme Ω (detto *spazio degli eventi* o *spazio dei campioni*) i cui elementi ω sono detti appunto *eventi elementari*. Questo insieme può essere finito, come nei casi più semplici visti sopra, oppure infinito: si prenda, ad esempio, una lampadina e si osservi per quanto tempo (in ore) resta accesa prima di bruciare; ogni numero reale non negativo può ritenersi un possibile risultato di questo esperimento.

Si chiama *evento* (casuale) ogni sottoinsieme di Ω (cioè ogni elemento di $\mathcal{P}(\Omega)$, l'insieme delle parti di Ω), compreso l'insieme vuoto $\emptyset$ (che viene detto *evento impossibile*) e Ω stesso (che viene detto *evento certo*, poiché, come abbiamo detto, uno degli elementi di Ω si verifica necessariamente). Così, nel lancio di un dado, l'insieme $\{2,4,6\}$ rappresenta l'evento (non elementare, ma composto di 3 eventi elementari) "uscita di un numero pari", l'insieme $\{1,2\}$ rappresenta l'evento "uscita

di un numero < 3" e così via. In questo modo, nella teoria della probabilità, gli insiemi vengono utilizzati come rappresentazione degli eventi, e tutto il corredo di simboli e di operazioni insiemistiche (esposto nella sez. 2) concorre a costruire la cosiddetta *algebra degli eventi*.

Qui sotto indichiamo brevemente l'interpretazione probabilistica delle principali operazioni insiemistiche.

Se $A, B, \dots$ sono eventi, cioè sottoinsiemi di Ω, si ha che:

$\omega \in A$	significa: l'evento elementare ω realizza l'evento A
$A \subset B$	significa: l'evento A implica l'evento B, cioè se si realizza A, si realizza B
$A \cup B$	significa: si verifica l'evento A oppure l'evento B
$A \cap B$	significa: si verifica sia l'evento A sia l'evento B
$A \cap B = \emptyset$	significa: i due eventi A e B non possono verificarsi entrambi, cioè sono incompatibili
$A \setminus B$	significa: si realizza A, ma non B
$C_\Omega A$	significa: non si realizza A (cioè si verifica uno degli eventi elementari che non stanno in A).

L'allievo è invitato a costruire esplicitamente degli esempi; è utile anche riesaminare le proprietà delle operazioni insiemistiche alla luce della interpretazione probabilistica.

Il passo successivo è l'introduzione del concetto di *probabilità* di un evento. Non è questo il luogo per discutere tale concetto; qui diremo solo che cosa si intende, dal punto di vista matematico, per probabilità. La questione presenta difficoltà ben diverse secondoché lo spazio degli eventi Ω è un insieme finito oppure infinito. Qui tratteremo solo il caso di Ω finito; qualche cenno al caso di Ω infinito sarà dato molto più avanti, quando tratteremo del problema della *misura* degli insiemi.

Sia dunque Ω finito, composto, diciamo, di n elementi; l'insieme delle parti $\mathcal{P}(\Omega)$ contiene dunque 2^n elementi, cioè ci sono 2^n eventi possibili.

Definizione 6.5 - *Diciamo probabilità in Ω una applicazione*

$$p\colon \mathcal{P}(\Omega) \to \mathbb{R}$$

(cioè una corrispondenza che associa ad ogni evento $A \in \mathcal{P}(\Omega)$ un numero reale $p(A)$, detto probabilità dell'evento A) che soddisfi le seguenti proprietà:

i) $p(A) \geq 0 \qquad \forall A \in \mathcal{P}(\Omega)$

ii) $p(\Omega) = 1$

iii) $p(A \cup B) = p(A) + p(B) \qquad \forall A, B \in \mathcal{P}(\Omega)$ con $A \cap B = \emptyset$.

Notiamo subito che dalle i) ... iii) seguono le seguenti:

iv) $p(CA) = 1 - p(A)$

v) $p(\emptyset) = 0$

vi) $0 \leq p(A) \leq 1 \qquad \forall A \in \mathcal{P}(\Omega)$

vii) $p(A \cup B) = p(A) + p(B) - p(A \cap B) \qquad \forall A, B \in \mathcal{P}(\Omega)$.

Infatti, scrivendo $A \cup CA = \Omega$ e usando la ii) e la iii) ricaviamo la iv). La v) è il caso particolare della iv) $A = \Omega$. La vi) proviene da i) e da iv). E infine la vii) si ottiene scrivendo $A \cup B = A \cup (B \setminus A \cap B)$ per cui, usando la iii)

$$p(A \cup B) = p(A) + p(B \setminus A \cap B) .$$

Osserviamo ora che $B = (A \cap B) \cup (B \setminus A \cap B)$ per cui

$$p(B) = p(A \cap B) + p(B \setminus A \cap B)$$

e, combinando questa con la precedente, otteniamo la vii).

Osservazione 6.2 - Usando il principio di inclusione-esclusione la vii) si generalizza subito al caso di n insiemi; la formula è la (6.4), dove, al posto di $|A|$, si legga $p(A)$.

***Esempio* 6.3** - Lancio del dado, $\Omega = \{1,2,3,4,5,6\}$. Possiamo ottemperare alle regole i), ii), iii) scegliendo

$$p(1) = p(2) = \ldots = p(6) = \frac{1}{6} .$$

Più in generale, se Ω è costituito da n eventi elementari, indicati con i numeri da 1 a n, $\Omega = \{1, 2, \ldots, n\}$, una scelta consistente di p è:

$$p(k) = \frac{1}{n} \qquad \forall k = 1, 2, \ldots, n . \tag{6.5}$$

A questo punto siamo in grado di calcolare la probabilità di qualunque dei 2^n eventi di Ω: se A è un evento composto di m eventi elementari, risulta:

$$p(A) = \frac{m}{n}$$

cioè il rapporto tra i casi "favorevoli" e i casi "possibili"; è questa la definizione elementare di probabilità data da Pascal.

Si sarà osservato naturalmente che la scelta fatta con le (6.5) non è l'unica possibile; infatti avremmo potuto porre

$$p(k) = p_k \qquad k = 1, 2, \ldots, n$$

con p_k numeri compresi tra 0 e 1 e tali che

$$\sum_{k=1}^{n} p_k = 1$$

(così da soddisfare la ii)) e per il resto arbitrari. Il sistema di assiomi i), ii), iii), che definisce la probabilità, è dunque consistente (nel senso che esistono oggetti reali che lo soddisfano) ma incompleto (nel senso che quegli assiomi non definiscono univocamente la funzione probabilità).

Se pensiamo al lancio del dado, è chiaro che noi saremo indotti alla scelta (6.5) se riterremo il dado non truccato e quindi se riteniamo "equo" assegnare ad ogni evento elementare la stessa probabilità; mentre, se il dado è truccato, la scelta da fare sarà diversa.

Saranno dunque solo l'esperienza, o considerazioni soggettive, a suggerire la scelta da operare: è su questo punto che le varie impostazioni (frequentista, soggettivista, etc.) del Calcolo della Probabilità si differenziano. Ma, per una discussione più approfondita sull'argomento, rimandiamo ai testi di probabilità.

***Esempio* 6.4** - Si lancino due dadi non truccati; calcolare la probabilità dei seguenti eventi:

A: il primo dado dà 4
B: il secondo dado dà 4
C: entrambi i dadi danno 4
D: almeno un dado dà 4
E: nessun dado dà 4
F: solo il primo dado dà 4
G: solo il secondo dado dà 4
H: solo un dado dà 4.

Lo spazio Ω è costituito dalle 36 coppie ordinate (i,j) con $i,j = 1,\dots,6$, rappresentanti gli eventi elementari, ad ognuno dei quali attribuiremo probabilità $1/36$.

L'evento A è costituito da 6 eventi elementari: le 6 coppie (i,j) con $i = 4$ (e $j = 1,2,\dots,6)$: perciò

$$p(A) = \frac{6}{36} = \frac{1}{6} ;$$

analogamente risulta

$$p(B) = \frac{1}{6} .$$

L'evento $C = A \cap B$ è costituito dalla sola coppia $(4,4)$ e perciò

$$p(C) = \frac{1}{36} .$$

L'evento D è $A \cup B$ e perciò

$$p(D) = p(A) + p(B) - p(A \cap B) = \frac{11}{36} .$$

L'evento E è $C_\Omega(A \cup B)$ e perciò

$$p(E) = 1 - p(D) = \frac{25}{36} .$$

L'evento F è $A \setminus B$: A contiene 6 elementi, B pure, $A \cap B$ un solo elemento; perciò $A \setminus B$ contiene $6 - 1 = 5$ elementi e pertanto

$$p(F) = \frac{5}{36} .$$

Analogamente $p(G) = 5/36$. Si osservi anche che $F \cap G = \emptyset$, cioè F e G si escludono a vicenda.

Perciò, essendo $H = F \cup G$, si ha:

$$p(H) = p(F) + p(G) = \frac{10}{36} .$$

Esercizi

1. Quanti numeri di 6 cifre si possono ottenere usando *ognuna* delle cifre 1, 2, 3, 4, 5, 6 esattamente una volta?

Quanti di questi numeri sono maggiori di 400'000?

2. In una squadra di calcio vi sono 4 difensori, 4 centrocampisti e 2 attaccanti. Quanti possibili schieramenti può adottare l'allenatore?

3. Quante diagonali vi sono in un poligono di n lati?

4. Quante sono le cinquine di interi naturali $(x_1, \ldots, x_5)$ che risolvono l'equazione

$$x_1 + x_2 + x_3 + x_4 + x_5 = 8?$$

E quelle che risolvono l'equazione

$$x_1 + x_2 + x_3 + x_4 + 3x_5 = 8?$$

5. Dare una dimostrazione di carattere combinatorio dell'uguaglianza

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

[Suggerimento: $\binom{n}{k}$ è il numero delle permutazioni di n oggetti di cui $n - k$ uguali ad a e k uguali a b. Il numero di tali permutazioni che hanno a come primo elemento è ..., ... che hanno b come primo elemento è ...].

6. Dimostrare che

$$\sum_{k=0}^{n} (-1)^k \binom{n}{k} = 0 .$$

7. Quante sono le schedine del totocalcio che contengono esattamente tre "2" e cinque "x"?

8. Sia **M** una matrice di ordine $m \times n$. Quanti minori principali di ordine k ($k \leq \min(m, n)$) si possono estrarre da **M**?

9. Quante sono le funzioni *iniettive* da $X = \{0, 1, 2\}$ in $Y = \{0, 1\}$? E da $X = \{0, 1, 2\}$ in $Y = \{0, 1, 2, 3\}$?

10. Risolvere i problemi 7 ed 8 enunciati nell'introduzione alla sezione.

11. Qual è la probabilità di ottenere un numero dispari nel lancio di 2 dadi?

12. Qual è la probabilità che tra n persone 2 siano nate nello stesso giorno? Trovare il minimo n per cui tale probabilità risulta $\geq 3/4$.

13. Un'urna contiene 1000 palline numerate da 0 a 999. Estraendo una pallina, qual è la probabilità che il numero abbia almeno una cifra pari?

7. INSIEMI INFINITI

Estendiamo l'ordinamento dei numeri cardinali, già introdotto sui cardinali finiti. Siano X, Y insiemi (eventualmente infiniti). Se esiste una applicazione *iniettiva* $f: X \to Y$ allora diremo che $\text{card}(X) \leq \text{card}(Y)$; scriveremo $\text{card}(X) < \text{card}(Y)$ se non è $\text{card}(X) = \text{card}(Y)$.
Ogni insieme finito di cardinalità n è in corrispondenza biunivoca con l'insieme campione $E_n = \{1, 2, \ldots, n\}$ che è un sottoinsieme proprio di $\mathbb{N}$. Chiaramente risulta $n \leq \text{card}(\mathbb{N})$, per ogni $n \in \mathbb{N}$. Mostriamo che è $n < \text{card}(\mathbb{N})$ per ogni n.

■ **Teorema 7.1** - $\mathbb{N}$ *è un insieme infinito.*

Dimostrazione - Se, per assurdo, fosse $\text{card}(\mathbb{N}) = n$, esisterebbe un'applicazione biiettiva $f: \mathbb{N} \to E_n$ la cui restrizione a E_n avrebbe come immagine un sottoinsieme proprio di E_n, diciamo E_m (con $m < n$). Ma allora tale restrizione determinerebbe una corrispondenza biunivoca tra insiemi, E_n e E_m, di cardinalità diversa. □

Dunque il numero cardinale $\text{card}(\mathbb{N})$ non è un intero naturale; si dice che è un *numero transfinito*. Si dice anche che l'insieme $\mathbb{N}$, come ogni altro insieme equipotente ad $\mathbb{N}$, ha la cardinalità o potenza del *numerabile* o, semplicemente, che è *un insieme numerabile*. Il nome è dovuto al fatto che, se A è numerabile, detta f una corrispondenza biunivoca con $\mathbb{N}$, $f: \mathbb{N} \leftrightarrow A$, potremo elencare (numerare) gli elementi di A in una successione, ponendo

$$a_n := f(n) .$$

***Esempio* 7.1** - L'insieme dei numeri pari è numerabile. Basta considerare la corrispondenza: $n \leftrightarrow 2n$.

L'esempio è interessante perché mostra che *un insieme infinito può essere messo in corrispondenza biunivoca con un suo sottoinsieme proprio* (cosa impossibile per gli insiemi finiti). Anzi, questa proprietà è caratteristica degli insiemi infiniti, sicché potrebbe essere utilizzata per definire questi insiemi.

Per la sua dimostrazione però si fa un uso essenziale dell'ultimo e più discusso assioma della teoria degli insiemi, l'*assioma della scelta* (v. Appendice). Utilizzando questo assioma si potrebbe dimostrare anche che ogni *insieme infinito contiene almeno un sottoinsieme numerabile* e perciò risulta che *non esistono insiemi infiniti aventi cardinalità strettamente minore della cardinalità del numerabile.*

Esaminiamo ora, dal punto di vista della cardinalità, il comportamento degli insiemi numerabili, rispetto a due operazioni insiemistiche fondamentali: l'unione e il prodotto (cartesiano).

▪ **Teorema 7.2** - *L'unione di una famiglia numerabile di insiemi numerabili è numerabile.*

Dimostrazione - Sia $A = \bigcup_{n=1}^{\infty} A_n$ (*), dove per ogni $n \geq 1$:

$$A_n = \{a_{n1}, a_{n2}, \ldots, a_{nm}, \ldots\} .$$

Possiamo allora costruire la seguente matrice (tabella a doppia entrata) in cui compaiono tutti e soli gli elementi di A:

Il procedimento illustrato in figura dalle frecce (detto procedimento diagonale di Cantor) indica in quale modo possono essere elencati gli elementi di A in una successione, saltando gli elementi che sono già comparsi una volta, nel caso che $A_n \cap A_m \neq \emptyset$ per qualche $m \neq n$. Dunque, A è numerabile. □

▪ **Teorema 7.3** - *Il prodotto $A_1 \times A_2 \times \ldots \times A_k$ di un numero finito di insiemi numerabili è numerabile.*

(*) Notazioni come $\bigcup_{n=0}^{\infty} A_n$ oppure $\bigcap_{n=0}^{\infty} B_n$ e simili sono usate al posto di $\bigcup_{n\in\mathbb{N}} A_n$ etc. Il simbolo ∞ si legge: infinito.

Dimostrazione - Per induzione. L'affermazione è vera per $k = 2$. Infatti gli elementi (a_{1n}, a_{2m}) del prodotto $A_1 \times A_2$ possono essere ordinati in una matrice scrivendo (a_{1n}, a_{2m}) all'incrocio della riga n-esima con la colonna m-esima; dopodiché si applica il procedimento diagonale. Supponiamo ora che l'insieme $B = A_1 \times A_2 \times \ldots \times A_{k-1}$ sia numerabile; allora $A_1 \times A_2 \times \ldots \times A_k = B \times A_k$ risulterà pure numerabile per quanto abbiamo sopra mostrato (prodotto con due soli fattori); ne consegue l'asserto. □

Notiamo in particolare che $\mathbb{N}^2 := \mathbb{N} \times \mathbb{N}$ *è numerabile.*

Le operazioni sopra illustrate consentono di costruire, a partire da insiemi numerabili, insiemi "molto più grandi", ma non "sostanzialmente più grandi", poiché non muta la cardinalità di questi insiemi.

Consideriamo invece l'operazione che fa passare da X a $\mathcal{P}(X)$.

■ **Teorema 7.4** - (di Cantor). *Per ogni insieme X risulta*

$$\operatorname{card}(X) < \operatorname{card}(\mathcal{P}(X)) .$$

Dimostrazione - Possiamo assumere X non vuoto (d'altra parte l'asserto è già noto per X finito). L'applicazione $X \ni x \mapsto \{x\} \in \mathcal{P}(X)$ è ovviamente iniettiva, da cui consegue che $\operatorname{card}(X) \leq \operatorname{card}(\mathcal{P}(X))$. Mostriamo che non è $\operatorname{card}(X) = \operatorname{card}(\mathcal{P}(X))$, cioè che non esiste alcuna applicazione biiettiva $f: X \to \mathcal{P}(X)$.

Per assurdo, esista una tale applicazione. Si consideri allora il sottoinsieme di X così definito:

$$Z := \{x \in X : x \notin f(x)\} .$$

Notiamo che $Z \neq \emptyset$ poiché evidentemente, se $\overline{x} = f^{-1}(\emptyset)$, $\overline{x} \notin f(\overline{x})$.

Dovrebbe esistere allora un certo $x_0 \in X$ del quale Z sarebbe l'immagine: $f(x_0) = Z$. Ma questa affermazione è contraddittoria, poiché, se $x_0 \notin f(x_0)$ allora $x_0 \in Z = f(x_0)$ e se $x_0 \in f(x_0)$ allora $x_0 \notin Z = f(x_0)$. □

Possiamo enunciare ora il paradosso di Cantor citato alla fine della sezione 2 relativo all'insieme Ω di tutti gli insiemi: esso non può essere un insieme, poiché risulterebbe $\operatorname{card}(\Omega) = \operatorname{card}(\mathcal{P}(\Omega))$. Contro il teorema 7.4.

Altri importanti esempi di insiemi infiniti numerabili e non si troveranno nel prossimo capitolo.

Appendice

CENNO ALLA TEORIA ASSIOMATICA DEGLI INSIEMI

1. Linguaggio della teoria

Col metodo assiomatico si formalizza la teoria degli insiemi fissandone rigorosamente il *linguaggio* e *l'apparato deduttivo.*

Tra le varie formalizzazioni presentiamo la prima, quella dovuta a E. Zermelo e A.A. Fraenkel.

Il linguaggio che si definisce è composto da:

- simboli di variabili: $x, y, z, \ldots, X, Y, Z, \ldots$
- simboli di costanti o elementi strutturalmente distinti
- simboli di punteggiatura e parentesi
- i connettivi proposizionali $\sim, \wedge, \vee, \Longrightarrow, \Longleftrightarrow$
- i quantificatori $\exists, \forall$
- i predicati:
  "insieme" (unario)
  " $=$ " (binario)
  " $\in$ " (binario) .

Si indicano poi quali sono le catene di simboli (o *formule*) corrette. Tali catene sono le *frasi ammissibili* nella teoria e sono definite *ricorsivamente* a partire dalle *formule atomiche* seguenti:

$x = y$; $x \in Y$; X è un insieme;

mediante l'uso ripetuto delle regole di formazione dei predicati che abbiamo visto nelle sezioni 1 e 2.

2. Apparato deduttivo

Consiste degli *assiomi* e delle *regole di deduzione.*

Le regole di deduzione sono quelle del calcolo dei predicati: essenzialmente il *modus ponens* e la seguente *regola di generalizzazione*: da $P(x)$ (con x libera) si deduce $\forall x \colon P(x)$. Questa regola è del resto di uso comune quando si deve dimostrare un teorema della forma $\forall x \colon P(x)$. Si fissa x (pensandolo arbitrariamente scelto) e si dimostra $P(x)$. La frase "per l'arbitrarietà di x segue la tesi" è legittimata dalla regola di generalizzazione.

Tra gli assiomi distinguiamo: quelli che appartengono al calcolo dei predicati e quelli relativi al predicato "=", su cui abbiamo già discusso; ad esempio quelli che legano i quantificatori.

Vi sono poi gli *assiomi specifici* della teoria, che qui di seguito elenchiamo. Essi formalizzano rigorosamente tutte le operazioni con gli insiemi che abbiamo trattato in 2.1.

ZF_1: *Assioma di estensione:* due insiemi che contengano gli stessi elementi sono uguali.

$$X = Y \Longleftrightarrow \forall z \colon (z \in X \Longleftrightarrow z \in Y) \, .$$

ZF_2: *Assioma dell'insieme vuoto:* esiste un insieme a cui non appartiene alcun

elemento: l'insieme vuoto, indicato con $\emptyset$:

$$\exists X\ \forall z : \sim (z \in X)\ .$$

Da ZF_1 e ZF_2 segue che l'insieme vuoto è unico.

ZF_3: *Assioma della coppia (non ordinata):* dati due insiemi x e y, esiste l'insieme Z i cui soli elementi sono x e y.

$$\forall x\ \forall y\ \exists Z\ \forall w : (w \in Z \iff (w = x \vee w = y))\ .$$

L'insieme costituito dai due elementi x ed y è indicato con $\{x, y\}$, coppia non ordinata. Si può poi definire la coppia *ordinata* (x, y) come la coppia (non ordinata) $\{x, \{x, y\}\}$. Si noti come (x, y) differisca da (y, x) che è definita come la coppia $\{y, \{x, y\}\}$.

ZF_4: *Assioma dell'unione:* se X è un insieme di insiemi, allora esiste un insieme S i cui elementi sono tutti e soli gli elementi di X.

$$\forall X\ \exists S \forall\ Z : (Z \in S \iff \exists W : (W \in X \wedge Z \in W))\ .$$

ZF_5: *Assioma di separazione o di specificazione:* per ogni insieme S ed ogni predicato p definito per ogni elemento di S, esiste un insieme (sottoinsieme di S) i cui elementi sono tutti e soli gli elementi di S che soddisfano p.

$$\forall S\ \exists Z\ \forall w : (w \in Z \iff w \in S \wedge p(w))\ .$$

Per indicare che A è un sottoinsieme di S, si scrive: $A \subseteq S$. Con questo assioma si definisce l'intersezione: basta considerare il predicato $p(x)$: "$x \in A$" (dove A è un insieme); allora l'insieme di cui l'assioma postula la esistenza è l'intersezione di S con A.

Analogamente, considerando il predicato: $x \notin A$, si ottiene il complementare di A rispetto a S.

ZF_6: *Assioma della potenza:* dato un insieme S esiste un insieme i cui elementi sono tutti e soli i sottoinsiemi di S; è l'insieme delle parti di S, $\mathcal{P}(S)$.

$$\forall S\ \exists T\ \forall w : (w \in T \iff w \subseteq S)\ .$$

Con i precedenti assiomi si possono definire tutte le operazioni insiemistiche, le *relazioni*, le *funzioni*; si possono costruire tutti gli insiemi finiti. L'esistenza di un insieme infinito è postulata nel seguente assioma.

ZF_7: *Assioma dell'infinito:* esiste un insieme che contiene l'insieme vuoto e contiene il successivo di ogni suo elemento (la definizione di successivo è stata data in 5.1).

$$\exists X : (\emptyset \in X \wedge \forall Y : (Y \in X \Longrightarrow \{Y, \{Y\}\} \in X))\ .$$

ZF_8: *Assioma di sostituzione:* data una funzione $f: S \to T$ esiste un insieme i cui elementi sono dati da $f(x)$ con $x \in S$. Tale insieme è l'immagine di f.

$$\forall f \; \exists Z \; \forall w : (w \in Z \iff \exists x : (x, w) \in f) \, .$$

Questo assioma consente di definire una funzione attraverso una proprietà caratteristica, anche senza conoscere l'insieme immagine.

ZF_9: *Assioma di regolarità o di fondazione:* sia p un predicato e supponiamo che esista un insieme con la proprietà p. Allora esiste un insieme X con la proprietà p e tale che nessuno dei suoi elementi ha la proprietà p.

$$(\exists Y : p(Y)) \Longrightarrow \exists X : (p(X) \wedge \forall z \in X : (\sim p(z))) \, .$$

ZF_{10}: *Assioma della scelta:* data una partizione S di un insieme T, esiste una funzione $f: S \to T$ (funzione di scelta) tale che, $\forall x \in S$, è $f(x) \in T$.

Questo assioma, oggetto di molte discussioni tra i matematici, afferma la possibilità di operare un numero infinito di "scelte" senza dover precisare il criterio di scelta. Esso è di fondamentale importanza per lo studio degli insiemi infiniti.

Insiemi numerici 2

Gli insiemi numerici di cui ci occuperemo sono:

$\mathbb{N}$ numeri interi naturali
$\mathbb{Z}$ numeri interi relativi
$\mathbb{Q}$ numeri razionali
$\mathbb{R}$ numeri reali
$\mathbb{C}$ numeri complessi

elencati in modo che ogni insieme sia un sottoinsieme proprio di quello scritto nella riga sottostante. Noi assumiamo che lo studente abbia già familiarità con questi insiemi (eccezion fatta per $\mathbb{C}$) almeno dal punto di vista algoritmico, cioè sappia operare con gli elementi di questi insiemi. Nostro obiettivo è quello di dare una definizione rigorosa di questi insiemi, a partire dal primo, l'insieme $\mathbb{N}$, di cui abbiamo già parlato nel precedente capitolo. Come vedremo, il passaggio da $\mathbb{N}$ a $\mathbb{Z}$ e da $\mathbb{Z}$ a $\mathbb{Q}$ non presenta difficoltà, come pure il passaggio da $\mathbb{R}$ a $\mathbb{C}$; il punto più delicato sarà la definizione di numero reale. Inoltre sarà nostro compito mettere in luce le più importanti proprietà strutturali di tali insiemi, mostrando come, nei successivi ampliamenti che da $\mathbb{N}$ conducono a $\mathbb{C}$, alcune proprietà, ritenute irrinunciabili, siano mantenute, altre lasciate cadere.

1. DA $\mathbb{N}$ A $\mathbb{Q}$

1.1 Rappresentazione dei numeri naturali

Nell'insieme $\mathbb{N}$ (le cui proprietà strutturali sono già state illustrate nel precedente capitolo) sono possibili, con alcune restrizioni, le operazioni elementari di addizione, sottrazione, moltiplicazione e divisione. La possibilità di eseguire agevolmente queste operazioni (problema di enorme importanza pratica) dipende però in larga misura dalla *rappresentazione* scelta per i numeri naturali. L'idea di una *rappresentazione posizionale* in base b, dove b è un intero > 1, (quella a cui siamo abituati è la rappresentazione posizionale in base 10) è fondata sul seguente teorema di aritmetica elementare (che il lettore potrà dimostrare per suo conto).

■ **Teorema 1.1** - *Se m, n sono numeri naturali e $n \geq m > 0$, allora esistono due soli numeri: q (quoziente) e r (resto) tali che*

$$n = mq + r \quad \text{con } 0 \leq r < m \; .$$

Vediamo ora come si può rappresentare un dato numero naturale $N > 0$ in base $b > 1$. Si comincia a dividere N per b trovando due soli numeri $q_1 \geq 0$ e $r_0\,(0 \leq r < b)$ tali che

$$N = q_1 b + r_0 \; ;$$

se $q_1 < b$ ci si ferma; se $q_1 \geq b$ si divide q_1 per b, trovando due soli numeri $q_2 \geq 0$ e $r_1\,(0 \leq r_1 < b)$ tali che

$$q_1 = q_2 b + r_1$$

e perciò:

$$N = q_2 b^2 + r_1 b + r_0 \; .$$

se $q_2 < b$ ci si ferma; se $q_2 \geq b$ si procede come prima. Dopo un numero *finito* di passi si trova $q_n = q_{n+1} b + r_n$ con $q_{n+1} = 0$ e $0 < r_n < b$; sarà allora:

$$\begin{aligned} N &= (((r_n b + r_{n-1})b + r_{n-2})b + \cdots + r_1)b + r_0 = \\ &= r_n b^n + r_{n-1} b^{n-1} + \cdots + r_1 b + r_0 \end{aligned}$$

dove tutti gli $r_i\,(i = 0, 1, \ldots, n)$ sono numeri $< b$.

Scegliendo allora b simboli (*cifre*) per i numeri minori di b possiamo individuare N con una sequenza ordinata di cifre, ponendo

$$N := (r_n r_{n-1} \ldots r_0)_b \; .$$

Questa è la rappresentazione di N in base b.

***Esempio* 1.1** - Se $b = 10$ e si scelgono i simboli

$$0, 1, 2, 3, 4, 5, 6, 7, 8, 9$$

per i numeri < 10, si ottiene la rappresentazione *decimale* dei numeri naturali, a cui siamo abituati. La sequenza di cifre $(1478)_{10}$ che scriviamo semplicemente 1478 sta ad indicare il numero

$$N = 1 \cdot 10^3 + 4 \cdot 10^2 + 7 \cdot 10 + 8 \; .$$

Se si scelgono $b = 2$ ed i simboli $0, 1$ per i numeri minori di 2, si ottiene la rappresentazione *binaria*. La sequenza di cifre $(1\;0\;0\;1\;0\;1\;0\;1)_2$ è la rappresentazione binaria di un naturale la cui rappresentazione decimale è

$$N = 2^7 + 2^4 + 2^2 + 1 = 149 \; .$$

In altri termini, $(1\ 0\ 0\ 1\ 0\ 1\ 0\ 1)_2$ e $(149)_{10}$ sono due rappresentazioni diverse dello stesso numero naturale.

1.2 I numeri interi relativi

I numeri interi relativi possono essere definiti formalmente a partire dai naturali nel modo seguente. Si consideri $\mathbb{N}^2 = \mathbb{N} \times \mathbb{N}$, (l'insieme delle coppie ordinate (m, n) di interi naturali) e si definisca un'*equivalenza* r ponendo

$$(m,n) \approx (m',n') \iff m + n' = m' + n \ .$$

Per esempio: $(5,3) \approx (8,6) \approx (2,0)\dots$ (verificare che si tratta proprio di un'equivalenza!). Quindi si consideri l'*insieme quoziente* $\mathbb{Z} = \mathbb{N}^2/r$ e sui suoi elementi, che indicheremo con $[(m,n)]$, si definiscano le operazioni di *somma* e *prodotto* con le regole seguenti:

$$\begin{aligned} &[(m,n)] + [(m',n')] := [(m+m', n+n')] \\ &[(m,n)] \cdot [(m',n')] := [(mm'+nn', nm'+mn')]. \end{aligned} \tag{1.1}$$

Osservazione 1.1 - Si noti che i simboli "+", "·" alla destra delle (1.1), si riferiscono a somma e prodotto tra naturali, mentre quelli a sinistra sono *definiti* dalle espressioni a secondo membro. Una simile situazione si presenterà più volte nel seguito.

(Verificare che, cambiando i rappresentanti nelle classi di equivalenza scritte a sinistra, non cambia la classe di equivalenza scritta a destra).

Si introduca poi un *ordinamento* (totale) ponendo

$$[(m,n)] \leq [(m',n')] \iff m + n' \leq m' + n \ .$$

Definizione 1.1 - *Gli elementi di* $\mathbb{Z}$ *così strutturato sono detti* **numeri interi relativi.**

In che senso $\mathbb{N}$ è contenuto in $\mathbb{Z}$? Nel senso che $\mathbb{N}$ è in corrispondenza biunivoca con il sottoinsieme di $\mathbb{Z}$ formato dalle (classi di equivalenza di) coppie del tipo $(m,0)$: $(m,0) \longleftrightarrow m$, e tale corrispondenza conserva le operazioni di somma e prodotto e la relazione d'ordine; possiamo dunque identificare $\mathbb{N}$ con l'insieme di tali coppie.

Infine semplifichiamo la notazione. Osserviamo che ogni coppia (m',n') è equivalente ad una del tipo

$$\begin{aligned} &(m,0) \ \text{ se } m' > n' \\ &(0,n) \ \text{ se } m' < n' \\ &(0,0) \ \text{ se } m' = n'. \end{aligned}$$

Scegliamo coppie di questo tipo come rappresentanti delle rispettive classi di equivalenza e scriviamo:

0 in luogo di $[(0,0)]$;
$+m$, o semplicemente m, in luogo di $[(m,0)]$;
$-n$ in luogo di $[(0,n)]$.

Risulta allora che, $\forall n,m: -m \leq 0 \leq n$; i numeri del tipo $-m$ sono detti *interi negativi*, quelli di tipo $+m$ *interi positivi*.

Ogni intero relativo p ammetterà una rappresentazione in base b del tipo:

$$p = \pm(r_n r_{n-1} \dots r_0)_b$$

con $r_0, r_1, \dots, r_n$ numeri naturali $< b$.

Lasciamo al lettore, che ne abbia voglia, la cura di verificare che valgono le usuali proprietà delle operazioni, le regole di calcolo, etc. È anche immediato constatare (sia direttamente che appellandosi al Teorema 1.7.3) che:

$\mathbb{Z}$ è un insieme numerabile.

1.3 I numeri razionali

In modo analogo si possono definire i numeri razionali a partire dagli interi relativi. Indichiamo il procedimento, lasciando al lettore la semplice verifica delle affermazioni.

Sia $\mathbb{Z}_0 = \mathbb{Z} \setminus \{0\}$ e consideriamo $\mathbb{Z} \times \mathbb{Z}_0$, cioè l'insieme delle coppie (p,q) di interi relativi con $q \neq 0$. Introduciamo una *equivalenza* r ponendo

$$(p,q) \approx (p',q') \iff pq' = p'q \,.$$

Per esempio $(2,5) \approx (6,15) \approx (-4,-10)\dots$ Quindi si passa al *quoziente* $\mathbb{Q} = (\mathbb{Z} \times \mathbb{Z}_0)/r$ e sugli elementi di questo si definiscono le operazioni di *somma* e *prodotto* con le regole seguenti:

$$[(p,q)] + [(p',q')] := [(pq' + p'q, qq')]$$
$$[(p,q)] \cdot [(p',q')] := [(pp', qq')] \,.$$

Si introduce un *ordinamento* totale ponendo

$$[(p,q)] \leq [(p',q')] \iff (pq' - p'q) \cdot qq' \leq 0 \,.$$

(I numeri $[(p,q)]$ con $pq < 0$ sono i negativi, $[(0,q)]$ è lo zero, e $[(p,q)]$ con $pq > 0$ sono i numeri positivi).

Infine si semplifica la notazione scrivendo

$$\frac{p}{q} \text{ in luogo di } [(p,q)] \,.$$

Definizione 1.2 - *Gli elementi di* $\mathbb{Q}$ *così strutturato si chiamano* **numeri razionali.**

Talvolta un numero razionale $[(p,q)]$ viene indicato semplicemente con una lettera, per esempio: a. Allora $-a$ indica il numero $[(-p,q)]$ e $1/a$ o a^{-1} il numero $[(q,p)]$ (se $p \neq 0$).

Si osserva che $\mathbb{Z}$ è in corrispondenza biunivoca con il sottoinsieme di $\mathbb{Q}$ formato dalle (classi di equivalenza di) coppie del tipo $(p,1)$: $(p,1) \leftrightarrow p$ e tale corrispondenza conserva le operazioni di somma e prodotto e la relazione d'ordine. In questo senso $\mathbb{Q}$ è un ampliamento di $\mathbb{Z}$. Infine, ricordando il teorema 1.7.3, risulta che:

$\mathbb{Q}$ è un insieme numerabile.

(Lo studente, per esercizio, dia di questa affermazione una prova diretta).

1.4 Struttura di $\mathbb{Q}$

Mettiamo ora in evidenza che l'insieme di numeri razionali sopra definito possiede una struttura di **corpo commutativo o campo ordinato.**

A questo scopo raggruppiamo le proprietà dei numeri razionali in tre gruppi: $\mathcal{R}_1$, $\mathcal{R}_2$, $\mathcal{R}_3$ qui sotto elencati.

$\mathcal{R}_1$. *È definita un'applicazione:* $\mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}$, *che si indica col segno "$+$" e si chiama addizione o somma, con le seguenti proprietà:*

1. $\forall a,b$: $a+b=b+a$ *(proprietà commutativa)*
2. $\forall a,b,c$: $(a+b)+c=a+(b+c)$ *(proprietà associativa)*
3. *esiste un elemento (elemento neutro), lo zero:* 0, *tale che:* $\forall a, a+0=a$
4. $\forall a$, *esiste un elemento (l'inverso di a rispetto alla somma),* $-a$, *tale che:* $a+(-a)=0$.

Si scrive poi $b-a$ in luogo di $b+(-a)$ e $a+b+c$ (senza mettere le parentesi) dal momento che non può esservi ambiguità. L'inverso di a (rispetto alla somma) si chiama *l'opposto* di a.

$\mathcal{R}_2$. *È definita un'applicazione:* $\mathbb{Q} \times \mathbb{Q} \to \mathbb{Q}$, *che si indica col segno "$\cdot$" e si chiama moltiplicazione o prodotto, con le seguenti proprietà:*

1. $\forall a,b$: $a \cdot b = b \cdot a$ *(proprietà commutativa)*
2. $\forall a,b,c$: $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ *(proprietà associativa)*
3. *esiste un elemento (elemento neutro), uno:* 1, *tale che:* $\forall a, a \cdot 1 = a$
4. $\forall a \neq 0$, *esiste un elemento (l'inverso di a rispetto al prodotto),* $1/a$, *indicato anche con* a^{-1}, *tale che:* $a \cdot a^{-1} = 1$.

Le operazioni di somma e prodotto sono legate dalla seguente proprietà

5. $\forall a,b,c$: $(a+b) \cdot c = a \cdot c + b \cdot c$ *(proprietà distributiva).*

Si scrive anche, come è noto, semplicemente ab invece di $a \cdot b$, abc (senza

parentesi) e così via. L'inverso di a (rispetto al prodotto) si chiama il *reciproco* di a.

Un insieme con le proprietà $\mathcal{R}_1$ e $\mathcal{R}_2$ si chiama *corpo commutativo* o *campo*. Da esse discende la possibilità di eseguire senza restrizione le quattro operazioni fondamentali: addizione e moltiplicazione (sopra definite) sottrazione (ponendo: $a - b = a + (-b)$) e divisione (ponendo $a/b = ab^{-1}$ purché $b \neq 0$).

La maggior parte delle proprietà dei razionali, ben note ad ogni studente, non sono proprietà particolari dei razionali, ma sono condivise da ogni altro corpo commutativo, in quanto derivano dalle proprietà $\mathcal{R}_1$ e $\mathcal{R}_2$. Le elenchiamo qui di seguito invitando ogni studente a svolgere le semplici dimostrazioni per esercizio.

1. Gli elementi neutri dell'addizione (0) e del prodotto (1) sono unici.
2. Gli elementi inversi di a rispettivamente nell'addizione $(-a)$ e nel prodotto $(a^{-1}, a \neq 0)$ sono unici.
3. $\forall a, b\colon a \cdot 0 = 0, -(-a) = a; (-a)(-b) = ab$
4. $\forall a, b$: $ab = 0 \Longrightarrow a = 0$ oppure $b = 0$ *(legge di annullamento del prodotto)*
5. $\forall a, b$ con $a \neq 0, b \neq 0$: $(ab)^{-1} = a^{-1}b^{-1}$.
6. $\forall a, b$ con $a \neq 0$ l'equazione $ax + b = 0$ ha un'unica soluzione $x = -ba^{-1}$.

$\mathcal{R}_3$. *È definita in* $\mathbb{Q}$ *una relazione di ordine totale* $<$, *compatibile con la struttura algebrica, cioè:*

1. $\forall a, b, c$: $a < b \Longrightarrow a + c < b + c$
2. $\forall a, b, c$ con $c > 0$: $a < b \Longrightarrow ac < bc$.

Un insieme con le proprietà $\mathcal{R}_1$, $\mathcal{R}_2$, $\mathcal{R}_3$ si dice *corpo commutativo ordinato* o *campo ordinato*. Valgono ancora le seguenti proprietà comuni ad ogni corpo commutativo ordinato (dimostrarle per esercizio):

1. $\forall a, b$: $a > 0 \Longrightarrow -a < 0$, $a > b \Longrightarrow a - b > 0$.
2. $\forall a, b, c$: $a < b$ e $c < 0 \Longrightarrow ac > bc$.
3. $\forall a\colon a \neq 0 \Longrightarrow a^2 > 0$; in particolare $1 > 0$.
4. L'equazione $x^2 + 1 = 0$ non ha soluzione in alcun corpo ordinato.
5. $\forall a, b$: $a > 0 \Longrightarrow a^{-1} > 0$, $a > b > 0 \Longrightarrow b^{-1} > a^{-1}$.

6. *Proprietà di densità:* $\forall a, b$ con $b > a$, esistono infiniti elementi maggiori di a e minori di b (se il campo è $\mathbb{Q}$ la cosa è evidente: basta prendere $c_0 = (a + b)/2$, e poi $c_1 = (a + c_0)/2$, $c_2 = (a + c_1)/2$ e così via).

Infine segnaliamo anche la seguente importante proprietà di $\mathbb{Q}$ (non valida in generale per ogni corpo ordinato):

7. *Proprietà di Archimede:* siano $a\,(= p/q)$ e $b\,(= r/s)$ numeri razionali positivi; allora esiste un intero naturale n tale che $na \geq b$ (basta prendere $n = rq$).

L'insieme dei razionali è un corpo ordinato ma non è l'unico; tuttavia esso è

il "più piccolo" tra i corpi ordinati nel senso precisato dal seguente teorema di algebra, di cui non riportiamo la dimostrazione.

■ **Teorema 1.2** - *Se S è un corpo ordinato, allora esiste un sottocorpo S_0 (*) che può essere identificato con $\mathbb{Q}$.*

1.5 Rappresentazione dei numeri razionali

La rappresentazione decimale dei numeri razionali è nota allo studente fin dalle scuole inferiori. Ci limitiamo qui a richiamare, senza dimostrazione, i risultati principali, avvertendo che, ciò che si dice qui per la base 10, può essere ripetuto, con ovvie modifiche, per una base b qualsiasi (b intero > 1).

Sia $a = p/q$ un razionale positivo, p e q naturali rappresentati in base 10. Dividendo p per q si trova

$$p = p_0 q + r_0 \qquad 0 \leq r_0 < q \ ;$$

e, successivamente, dividendo $10r_0$ per q:

$$10r_0 = \alpha_1 q + r_1 \qquad 0 \leq r_1 < q \ ;$$

dividendo $10r_1$ per q:

$$10r_1 = \alpha_2 q + r_2 \qquad 0 \leq r_2 < q \ ,$$

$$\vdots$$

dove p_0 è un intero non negativo e $\alpha_1, \alpha_2, \ldots$ sono le cifre decimali. Si scrive

$$a = p_0 . \alpha_1 \alpha_2 \cdots \alpha_n \cdots$$

e risulta, per ogni n,

$$p_0 . \alpha_1 \alpha_2 \cdots \alpha_n \leq a < p_0 . \alpha_1 \alpha_2 \cdots \alpha_n + \frac{1}{10^n} \ .$$

La successione $p_0, \alpha_1, \alpha_2, \ldots, \alpha_n, \ldots$ viene detta *allineamento decimale* (p_0 vien detto *parte intera* di a e indicato con $[a]$).

È noto (lo studente ritrovi la dimostrazione) che dividendo due interi (positivi) si può ottenere soltanto un **allineamento periodico**, cioè, da un certo posto in poi,

(*) Se S è un corpo, per sottocorpo di S si intende un sottoinsieme di S che costituisca un corpo rispetto alle stesse operazioni di S.

un blocco di un numero finito di cifre, detto *periodo*, si ripete indefinitamente. Per esempio:

$$\frac{2}{5} = 0.4000\cdots$$

$$\frac{107}{249} = 0.4323232\cdots$$

Se il periodo è 0 l'allineamento si dice **limitato** e il periodo non si scrive ($2/5 = 0.4$); diversamente, l'allineamento si dice **illimitato periodico** e si scrive soltanto il primo periodo con un soprassegno ($107/249 = 0.4\overline{32}$).

È noto che non si può ottenere un allineamento con periodo 9. Infatti, se il numero razionale a ammettesse una rappresentazione della forma

$$p_o.\alpha_1\cdots\alpha_k\alpha_{k+1}\cdots \quad \text{con } 9 = \alpha_{k+1} = \alpha_{k+2} = \ldots$$

si avrebbe, per ogni $n \geq k$,

$$p_0.\alpha_1\cdots\alpha_k\cdots\alpha_n = p_0.\alpha_1\cdots\alpha_k + \frac{9}{10^{k+1}} + \cdots + \frac{9}{10^n} =$$

(ricordando la formula 1.(5.1) per la somma dei termini in progressione geometrica)

$$= p_0.\alpha_1\cdots\alpha_k + \frac{9}{10^{k+1}}\frac{1-\dfrac{1}{10^{n-k}}}{1-\dfrac{1}{10}} = p_0.\alpha_1\cdots\alpha_k + \frac{1}{10^k} - \frac{1}{10^n}\ .$$

e dovrebbe essere

$$p_0.\alpha_1\cdots\alpha_k + \frac{1}{10^k} - \frac{1}{10^n} \leq a < p_0.\alpha_1\cdots\alpha_k + \frac{1}{10^k}$$

Ma la disuguaglianza di sinistra, dovendo valere per ogni $n \geq k$, implica $p_0.\alpha_1\cdots\alpha_k + 1/10^k \leq a$, in contraddizione con la disuguaglianza di destra.

Tuttavia è comodo ammettere anche tali allineamenti, con la convenzione che: l'allineamento illimitato $p_0.\alpha_1\alpha_2\cdots\alpha_k 9\ 9\ 9\cdots$ viene identificato con l'allineamento limitato $p_0.\alpha_1\alpha_2\cdots(\alpha_k+1)$. Per esempio, $0.\overline{9}$ è identificato con 1. Ciò è consistente con quanto sopra detto, poiché, se facessimo corrispondere agli allineamenti $p_0.\alpha_1\alpha_2\cdots(\alpha_k+1)$ e $p_0.\alpha_1\alpha_2\cdots\alpha_k\overline{9}$ due numeri a_1 e a_2 rispettivamente, con $a_1 > a_2$, dovrebbe essere, per ogni intero $m \geq 0$, $0 < a_1 - a_2 < 10^{-(m+k)}$, una contraddizione.

In questo modo, ad ogni allineamento decimale periodico, resta univocamente associato un numero razionale (positivo). Per esempio:

$$4.25 \to \frac{17}{4} \qquad 2.31\overline{7} \to \frac{231}{100} + \frac{7}{900}\ .$$

Ora, le cose dette per i razionali positivi si applicano anche a quelli negativi, che saranno rappresentati da un allineamento preceduto dal segno $-$. Gli allineamenti $\pm 0.000\ldots$ saranno ovviamente identificati con un unico allineamento rappresentato semplicemente con 0.

Lo studente saprà da solo trasportare sugli allineamenti decimali periodici con segno le operazioni e l'ordinamento esistenti in $\mathbb{Q}$.

A questo punto *l'insieme degli allineamenti decimali periodici con segno (con le convenzioni fatte) è un campo ordinato in corrispondenza biunivoca con il campo ordinato dei razionali e tale corrispondenza conserva le operazioni e l'ordinamento*. Si dice che $\mathbb{Q}$ è *isomorfo* all'insieme degli allineamenti periodici e perciò possiamo d'ora in avanti *identificare* i numeri razionali con tali allineamenti.

Un'altra *rappresentazione*, stavolta di tipo *geometrico*, di $\mathbb{Q}$ si può ottenere associando ad ogni numero razionale un punto della retta euclidea. Ad un punto di questa, scelto arbitrariamente, si associa 0 e ad un altro, distinto dal primo, si associa 1, individuando così il segmento orientato 01 che costituisce *l'unità di misura*. A questo punto si ha una corrispondenza biunivoca tra i numeri razionali e quei punti P della retta che sono estremi dei segmenti orientati OP *misurabili* o *commensurabili* (nel senso classico della geometria elementare, cioè che hanno un multiplo comune) con 01.

Esercizi

1. Siano $n_1 = 961$, $n_2 = 93$, $n_3 = 100$ numeri espressi in base 10. Trovare l'espressione di n_1, n_2, n_3 in base $2, 8, 16$.

2. Trovare l'espressione decimale dei numeri seguenti

$$n_1 = (110101)_2 \ , \ n_2 = (4321)_5 \ , \ n_3 = (AB10)_{16}$$

(in base $16, A$ è il simbolo per il 10 decimale, B per l'11).

3. Scrivere la rappresentazione decimale delle seguenti frazioni:

$$\frac{8}{15} \ , \ \frac{121}{7} \ , \ \frac{1}{125} \ .$$

4. Rappresentare come frazioni i seguenti numeri razionali

$$8.9\overline{3} \ , \ 0.821\overline{4} \ , \ 7.5\overline{463} \ .$$

5. Trovare un modo esplicito per enumerare i razionali.
[Suggerimento: se $r = m/n$ è razionale positivo, m, n primi tra loro, definire *altezza* di r il numero intero $m + n$. Porre $r_0 = 0$ ed enumerare i razionali per altezze crescenti ...].

6. Dimostrare le proprietà 1-6 di pagina 74.

7. Dimostrare le proprietà 1-6 dei campi ordinati a pagina 74.

2. I NUMERI REALI

La struttura di campo ordinato dei razionali assolve alla maggior parte degli scopi pratici del calcolo, nel senso che si può esprimere con un numero razionale la misura di ogni grandezza con sufficiente precisione. Tuttavia abbiamo già notato (v. 1.1.4) che esistono grandezze geometriche che non sono commensurabili con l'unità di misura adottata, qualunque essa sia, come la diagonale del quadrato il cui lato ha lunghezza unitaria. In altre parole, pensando alla rappresentazione sulla retta euclidea dei numeri razionali, si vede che esistono sulla retta ancora molti posti vuoti. Sorge spontanea la domanda: è possibile ampliare l'insieme dei razionali in modo da avere ancora un campo ordinato, i cui elementi (numeri) permettano di esprimere *tutte* le grandezze geometriche? D'altronde, l'idea dell'ampliamento si presenta spontanea anche se si considera la rappresentazione algebrica dei razionali, come allineamenti decimali periodici: è possibile strutturare come campo ordinato l'insieme di *tutti* gli allineamenti decimali?
A tali domande daremo risposta in questo paragrafo, costruendo appunto il campo ordinato dei numeri reali, il quale contiene quello dei razionali come suo sottocampo ed ha in più la proprietà di completezza ($\mathcal{R}_4$) che ampiamente illustreremo.
Noi daremo, del numero reale, una definizione basata sul concetto di allineamento decimale. Perché questa scelta, invece di tentare una definizione formale del tipo di quelle che ci hanno permesso di passare da $\mathbb{N}$ a $\mathbb{Z}$ e da $\mathbb{Z}$ a $\mathbb{Q}$?
Per due principali ragioni: innanzitutto una definizione di questo tipo non è così semplice come quelle già viste, perché non basta utilizzare coppie (o terne, o n-uple) di numeri razionali per costruire i numeri reali, bensì bisognerebbe considerare *successioni*; noi riteniamo che questo procedimento potrà essere ben compreso e apprezzato dallo studente solo più avanti, dopo che egli si sia familiarizzato con il concetto di spazio metrico completo; e infatti vi ritorneremo.
La seconda ragione è che questa presentazione ci è sembrata la più semplice e la più aderente all'idea che ogni studente si fa di un numero reale; è questo che ce l'ha fatta preferire ad altre definizioni pure "costruttive", come, ad esempio, le *sezioni di Dedekind* o le *coppie di classi contigue*. Infine, una definizione costruttiva è stata preferita ad una assiomatica perché ci è sembrato, secondo la nostra esperienza, che quest'ultima lasci lo studente (ancora inesperto dei metodi astratti della matematica) piuttosto scettico e talvolta addirittura incredulo di fronte ad una tale impostazione.

2.1 Definizione di numero reale

Definizione 2.1 - *Definiamo* **numero reale** *un allineamento decimale con segno. Se l'allineamento è periodico (limitato o no) il numero si dirà* **razionale**, *altrimenti si dirà* **irrazionale**. L'insieme dei numeri reali sarà indicato con $\mathbb{R}$.

Una definizione più formale è la seguente: *per numero reale si intende una terna* (σ, p, φ) dove:

σ è il segno ($+$ o $-$),

p è un intero naturale: $p \in \mathbb{N}$,
φ è una successione di cifre decimali, cioè un'applicazione da $\mathbb{N}_+$ nell'insieme $\{0,1,2,3,4,5,6,7,8,9\}$ (*)

Naturalmente questa è una definizione solo formale; se infatti vogliamo che i nuovi enti formino un *campo ordinato* dobbiamo definire per essi le operazioni previste in tale struttura e introdurre una relazione d'ordine compatibile con essa, in modo tale che tutto sia coerente con quanto già definito nell'ambito dei razionali.
Intanto cominciamo col definire *uguali* due numeri reali $a = (\sigma_1, p_1, \varphi_1)$ e $b = (\sigma_2, p_2, \varphi_2)$ se e solo se:

$$\sigma_1 = \sigma_2 \quad , \quad p_1 = p_2 \quad , \quad \varphi_1 = \varphi_2 \, .$$

Le convenzioni già fatte per gli allineamenti periodici (circa quelli di periodo 9 etc.) restano in vigore.
Chiameremo *negativi* gli allineamenti col segno $-$ e *positivi* quelli col segno $+$.
L'insieme dei numeri reali positivi (negativi) sarà indicato con $\mathbb{R}_+(\mathbb{R}_-)$.

2.2 Ordinamento

Introduciamo sugli allineamenti l'ordinamento lessicografico, cioè:

1. Se a e b sono positivi, $a \neq b$, così rappresentati

$$a = +p.\alpha_1\alpha_2\cdots\alpha_n\cdots$$

$$b = +q.\beta_1\beta_2\cdots\beta_n\cdots$$

diremo che

$$a > b$$

se $p > q$ oppure $p = q$ ed esiste un intero positivo m tale che $\alpha_m > \beta_m$ e, se $m > 1$, $\alpha_k = \beta_k$ per $k \leq m-1$.
Perciò se a è positivo, risulta $a > 0$.

2. Se a è negativo e b positivo o nullo, poniamo per definizione: $a < b$. In particolare risulta $a < 0$ se a è negativo.

Restano da ordinare i numeri negativi. Cominciamo col porre, se $a = \pm p.\alpha_1\alpha_2\cdots\alpha_n\cdots$:

$$-a := \mp p.\alpha_1\alpha_2\cdots\alpha_n\cdots$$
$$|a| := +p.\alpha_1\alpha_2\cdots\alpha_n\cdots \; ;$$

(*) $\mathbb{N}_+$ indica l'insieme degli interi positivi: $\mathbb{N}_+ := \mathbb{N}\setminus\{0\}$.

$|a|$ si dice **valore assoluto** o **modulo** di a. Chiaramente si ha:

$$|a| = \begin{matrix} a & \text{se } a > 0 \\ -a & \text{se } a \leq 0 \end{matrix}$$

3. Se $a < 0$, $b < 0$ porremo allora

$$a < b \Longleftrightarrow |b| < |a| .$$

Lasciamo al lettore la facile verifica che la relazione così introdotta tra i numeri reali soddisfa le 3 proprietà degli ordinamenti.

L'insieme dei numeri reali resta così *totalmente ordinato*; possiamo perciò parlare di insiemi (di numeri reali) limitati inferiormente e/o superiormente, di massimo e minimo, di estremo superiore e inferiore (si rivedano, a questo proposito, le definizioni date in 1.3.4).

2.3 Struttura algebrica

Introduciamo ora le operazioni aritmetiche. Sarà essenziale il concetto di **successione stabilizzata** di numeri reali (non negativi) che ora definiamo.

Cominciamo col dire che: *se* $\{p_n\}$ *è una successione di numeri naturali ed esiste un intero naturale* N *tale che per* $n \geq N$ *risulta* $p_n = p$ (costante) *diremo che la successione* $\{p_n\}$ **è definitivamente costante, o stabilizzata.**

***Esempio* 2.1** - Una successione $\{p_n\}$ di interi, non decrescente (cioè $p_n \leq p_{n+1}$, $\forall n \in \mathbb{N}$) e limitata superiormente è definitivamente costante.

Sia ora $\{a_n\}$ una successione di numeri reali non negativi:

$$\begin{aligned} a_1 &= \alpha_{10} \,.\, \alpha_{11} \;\; \alpha_{12} \;\; \alpha_{13} \cdots \\ a_2 &= \alpha_{20} \,.\, \alpha_{21} \;\; \alpha_{22} \;\; \alpha_{23} \cdots \\ a_3 &= \alpha_{30} \,.\, \alpha_{31} \;\; \alpha_{32} \;\; \alpha_{33} \cdots \\ &\vdots \end{aligned} \tag{2.1}$$

(seguendo l'uso, non scriviamo il segno $+$). Questi allineamenti costituiscono un quadro a doppia entrata (una matrice) infinita. Consideriamo le colonne di questo quadro: nella colonna di posto 0 figurano le parti intere α_{n0} dei numeri a_n, nella colonna di posto $k \geq 1$ figurano le k-esime cifre decimali α_{nk} dei numeri a_n.

Definizione 2.2 - *Se, per ogni* $k \geq 0$, *la sequenza di interi* α_{nk} *è definitivamente costante, la successione di reali* $\{a_n\}$ *si dice* **stabilizzata.**

Osserviamo che, *se* $\{a_n\}$ *è stabilizzata, esiste un unico numero reale*

$$a := \gamma_0 \,.\, \gamma_1 \gamma_2 \gamma_3 \cdots$$

tale che, per ogni $k \geq 0$, risulti $\gamma_k = \alpha_{nk}$ per n sufficientemente grande; scriveremo allora

$$a_n \rightrightarrows a$$

(si legge: la successione $\{a_n\}$ individua a).

Esempi

2.2. La successione

$$\begin{array}{l} 0.10101010\cdots \\ 0.01010101\cdots \\ 0.10101010\cdots \\ 0.01010101\cdots \\ \vdots \end{array}$$

non è stabilizzata.

2.3. La successione

$$\begin{array}{l} 0.1234567\cdots \\ 0.1123456\cdots \\ 0.1122345\cdots \\ 0.1122334\cdots \end{array}$$

è stabilizzata e individua il numero $a = 0.1122334455\ldots$.

Il seguente lemma risulta fondamentale.

Lemma 2.1 - *Sia $\{a_n\}$ una successione di numeri reali non negativi, non decrescente* (cioè $a_n \leq a_{n+1}, \forall n \in \mathbb{N}$) *e limitata superiormente* (da un intero M, per esempio). *Allora $\{a_n\}$ è stabilizzata, cioè $a_n \rightrightarrows a$, dove a soddisfa le limitazioni*

$$a_n \leq a \leq M \ .$$

Dimostrazione - Sotto le condizioni del lemma, gli interi che costituiscono la colonna zero della matrice (2.1), cioè gli interi $\{\alpha_{n0}\}$, costituiscono una successione non decrescente e limitata superiormente da M e perciò definitivamente costante; dunque esiste un intero γ_0 per cui risulta $\alpha_{n0} = \gamma_0$ per $n \geq n_0$ sufficientemente grande, con $\gamma_0 \leq M$.

Si procede ora per induzione. Supponiamo che siano stabilizzate le cifre fino alla k-esima, cioè, che per $n \geq n_k$ si abbia

$$a_n = \gamma_0 \cdot \gamma_1 \gamma_2 \cdots \gamma_k \alpha_{n,k+1} \alpha_{n,k+2} \cdots$$

e che inoltre,

$$\gamma_0 \cdot \gamma_1 \gamma_2 \cdots \gamma_k \leq M \ .$$

Poiché $\{\alpha_{n,k+1}\}$ è una successione di cifre tra 0 e 9, non decrescente (per $n \geq n_k$) deve essere definitivamente costante; quindi esiste n_{k+1} tale che, per $n \geq n_{k+1}$ si ha

$$a_n = \gamma_0 \cdot \gamma_1\gamma_2 \cdots \gamma_k\gamma_{k+1}\alpha_{n,k+2}\alpha_{n,k+3} \cdots \leq M$$

Quindi anche

$$\gamma_0 \cdot \gamma_1\gamma_2 \cdots \gamma_k\gamma_{k+1} \leq M \ .$$

Per induzione otteniamo allora che $a_n \rightrightarrows a := \gamma_0 \cdot \gamma_1\gamma_2\gamma_3 \cdots$, con $a \leq M$.

Proviamo ora che $a_n \leq a$ per ogni n. Se, per assurdo, fosse $a_{\overline{n}} > a$ per un certo $\overline{n}$, sarebbe, supponendo che le prime k cifre di $a_{\overline{n}}$ siano stabilizzate:

$$a_{\overline{n}} = \gamma_0 \cdot \gamma_1\gamma_2 \cdots \gamma_k\alpha_{\overline{n},k+1}\alpha_{\overline{n},k+2} \cdots > a := \gamma_0 \cdot \gamma_1\gamma_2 \cdots \gamma_k\gamma_{k+1} \cdots \ .$$

Questo implica che $\alpha_{\overline{n},k+1} > \gamma_{k+1}$; ma ciò contraddice il fatto che per $m \geq \overline{n}$, $\{\alpha_{n,k+1}\}$ è non decrescente. □

Possiamo ora definire le operazioni aritmetiche. Siano

$$a = p\,.\,\alpha_1\alpha_2\alpha_3 \cdots \qquad b = q\,.\,\beta_1\beta_2\beta_3 \cdots$$

due numeri reali non negativi. Introduciamo i numeri troncati (razionali)

$$a^{(n)} = p\,.\,\alpha_1\alpha_2 \cdots \alpha_n 00 \cdots$$
$$b^{(n)} = q\,.\,\beta_1\beta_2 \cdots \beta_n 00 \cdots \ .$$

Osserviamo ora che le successioni di razionali

$$\{a^{(n)} + b^{(n)}\} \qquad \{(a^{(n)} \cdot b^{(n)})^{(n)}\}$$

(dove $+$ e $\cdot$ indicano somma e prodotto di razionali!) sono *non decrescenti e limitate superiormente* (da $p + q + 2$ e da $(p + 1)(q + 1)$ rispettivamente); dunque, per il lemma 2.1, sono *stabilizzate*. Possiamo allora definire (per $a \geq 0, b \geq 0$):

somma: *$a + b$ è il numero reale s tale che*

$$a^{(n)} + b^{(n)} \rightrightarrows s$$

prodotto: *$a \cdot b$ è il numero reale p tale che*

$$(a^{(n)} \cdot b^{(n)})^{(n)} \rightrightarrows p \ .$$

Osserviamo ora che le successioni di razionali

$$\{a^{(n)} - (b^{(n)} + 10^{-n})\} \qquad \left\{\left(\frac{a^{(n)}}{b^{(n)} + 10^{-n}}\right)^{(n)}\right\}$$

sono non decrescenti, poiché $\{b^{(n)} + 10^{-n}\}$ è non crescente, e limitate superiormente (rispettivamente da $p + 1$ e da $(p + 1)/b^{(n_0)}$, dove $b^{(n_0)}$ è una troncata di b, positiva).

Definiamo:

differenza: $a - b$ $(a > b)$ *è il numero reale d tale che*

$a^{(n)} - (b^{(n)} + 10^{-n}) \rightrightarrows d$ (*)

quoziente: a/b (con $b > 0$) *è il numero reale q tale che*

$$\left(\frac{a^{(n)}}{b^{(n)} + 10^{-n}}\right)^{(n)} \rightrightarrows q \, .$$

Le operazioni così definite si estendono in modo naturale ai numeri reali di segno qualunque. Ad esempio:

se $a < 0, b < 0$ si pone

$$a + b := -(|a| + |b|)$$
$$a \cdot b := |a| \cdot |b| \, ;$$

se $a > 0, b < 0$ si pone

$$a + b := \begin{cases} a - |b| & \text{se } a \geq |b| \\ -(|b| - a) & \text{se } a \leq |b| \end{cases}$$
$$a \cdot b := -|a| \cdot |b| .$$

▪ **Teorema 2.2** - *L'insieme dei numeri reali è un campo (corpo commutativo) ordinato rispetto alla relazione d'ordine e alle operazioni di somma e prodotto definite sopra.*

La dimostrazione di questo teorema è una noiosa verifica delle proprietà $\mathcal{R}_1$, $\mathcal{R}_2$, $\mathcal{R}_3$ dei campi ordinati; essa viene riportata in Appendice alla fine della sezione. Come conseguenza, valgono, per i numeri reali, tutte le proprietà dei campi ordinati elencate in 1.4. I numeri razionali risultano così un sottocorpo del corpo dei reali.

2.4 Proprietà di completezza

Veniamo ora alla proprietà fondamentale che distingue il corpo dei reali da quello dei razionali: la **completezza**.

Riprendiamo il concetto di estremo superiore (inferiore) già introdotto per ogni insieme ordinato. Sia $A \subset \mathbb{R}$ (non vuoto) limitato superiormente (inferiormente). Il $\sup A$ ($\inf A$) è il minimo (massimo) dei maggioranti (minoranti) di A. Esso coincide col massimo (minimo) di A quando questo esista.

(*) n va scelto abbastanza grande in modo che $a^{(n)} - (b^{(n)} + 10^{-n}) \geq 0$.

Una definizione equivalente è la seguente: *si chiama* **estremo superiore (inferiore)** *di A il numero reale L (l) tale che*:

i) $\forall x \in A: x \leq L \ (x \geq l)$

ii) $\forall \varepsilon > 0, \ \exists x \in A: L - \varepsilon < x \quad (x < l + \varepsilon)$.

La i) dice infatti che L (l) è un maggiorante (minorante) di A; la ii) dice che non esiste un maggiorante (minorante) più piccolo (grande) di L (l).

È noto che per insiemi ordinati qualsivoglia, queste quantità possono anche mancare.

***Esempio* 2.4** - Sia $A = \{x \in \mathbb{Q}: x > 0, x^2 < 2\}$ (*) .

A è limitato (sia inferiormente che superiormente), possiede estremo inferiore (0), ma non possiede minimo, non possiede massimo e neanche estremo superiore (provarlo per esercizio).

Ma vedremo che l'insieme:

$$B = \{x \in \mathbb{R}: x > 0, x^2 < 2\}$$

ancora non possiede né massimo né minimo, ma possiede sia estremo inferiore (0) che estremo superiore $(\sqrt{2})$.

La circostanza messa in luce dal precedente esempio non è casuale, come mostra il seguente fondamentale teorema.

Introduciamo le seguenti notazioni: dati due numeri reali a, b, con $a < b$, poniamo:

$$[a,b] := \{x \in \mathbb{R}: a \leq x \leq b\}$$
$$[a,b) := \{x \in \mathbb{R}: a \leq x < b\}$$
$$(a,b) := \{x \in \mathbb{R}: a < x < b\}$$
$$(a,b] := \{x \in \mathbb{R}: a < x \leq b\} \ (^{**}) .$$

Tutti questi insiemi si chiamano **intervalli** (limitati); si chiamano intervalli (illimitati) anche le semirette, cioè insiemi del tipo

$$\{x \in \mathbb{R}: x \geq a\} \ , \ \{x \in \mathbb{R}: x > a\} \ , \ \{x \in \mathbb{R}: x \leq b\} \ , \ \{x \in \mathbb{R}: x < b\} \ .$$

■ **Teorema 2.3** - (di completezza). *Sia $A \subset \mathbb{R}$, $A \neq \emptyset$.*

(*) A rigore dovremmo scrivere: $\{x \in \mathbb{Q}: x > 0 \text{ e } x^2 < 2\}$; ma spesso la virgola sostituirà il connettivo "e".

(**) Sarà sempre chiaro dal contesto quando la notazione (a, b) indicherà un intervallo oppure una coppia ordinata.

$\mathcal{R}_4$. *Se A è limitato superiormente (inferiormente), A possiede estremo superiore (inferiore).*

Dimostrazione - Diamo la dimostrazione solo per l'estremo superiore.

Se A è finito, esiste il massimo di A e allora l'estremo superiore è il massimo. Sia dunque A infinito.

Poiché A è limitato superiormente, esiste $y_0 \in \mathbf{R}$ tale che: $y_0 \geq x, \forall x \in A$. Sia $x_0 \in A$, $x_0 < y_0$ e consideriamo l'intervallo $[x_0, y_0]$. Dividiamo a metà questo intervallo ottenendo i due intervalli $[x_0, c]$ e $[c, y_0]$ dove $c = (x_0 + y_0)/2$.

Scegliamo dei due intervalli uno in cui ci siano elementi di A, optando per il secondo se ci sono elementi di A in entrambi. Indichiamo tale intervallo con $[x_1, y_1]$. Evidentemente risulta $y_1 \geq x$, $\forall x \in A$ e

$$y_1 - x_1 = \frac{y_0 - x_0}{2} \ .$$

Procediamo ora con $[x_1, y_1]$ come abbiamo fatto con $[x_0, y_0]$ e così di seguito. Otteniamo una sequenza di intervalli $[x_n, y_n]$ con queste proprietà:

1) $[x_0, y_0] \supset [x_1, y_1] \supset \ldots \supset [x_n, y_n] \supset [x_{n+1}, y_{n+1}] \supset \ldots$

2) $y_n - x_n = (y_0 - x_0)/2^n$ e $[x_n, y_n]$ contiene punti di A

3) $y_n \geq x \ \forall x \in A$ e per ogni n

4) $\{x_n\}$ è non decrescente e limitata superiormente (da y_0) e quindi esiste un numero $\overline{x}$ per cui $x_n \rightrightarrows \overline{x}$. Inoltre, per ogni k ed ogni n, $x_n < y_k$ e quindi $x_n \leq \overline{x} \leq y_k$ per ogni k ed ogni n. In particolare, per ogni n,

$$x_n \leq \overline{x} \leq y_n \ . \tag{2.2}$$

Individuato così il numero $\overline{x}$ dimostriamo che è: $\overline{x} = \sup A$. Basta provare che $\overline{x}$ verifica le proprietà i) e ii) della definizione di estremo superiore.

i) $\overline{x} \geq x, \forall x \in A$. Per assurdo: esista $y \in A$ con $\overline{x} < y$. Poniamo $r = y - \overline{x}$ e scegliamo $\overline{n}$ tale che

$$\frac{y_0 - x_0}{2^{\overline{n}}} < r \ .$$

Allora avremmo, per la (2.2) e la proprietà 2):

$$y_{\overline{n}} - \overline{x} \leq y_{\overline{n}} - x_{\overline{n}} = \frac{y_0 - x_0}{2^{\overline{n}}} < r = y - \overline{x}$$

e cioè $y_{\overline{n}} < y$; ciò contraddice la 3).

ii) $\forall \varepsilon > 0$, $\exists y \in A : y > \overline{x} - \varepsilon$. Ancora per assurdo: fissiamo $\varepsilon > 0$ e supponiamo che $\forall y \in A$ risulti $y \leq \overline{x} - \varepsilon$. Poiché $\overline{x} - \varepsilon < \overline{x} \leq y_n$ per ogni n, si ha che l'intervallo $[\overline{x} - \varepsilon, y_n]$ non contiene punti di A. Scegliamo $\overline{n}$ tale che $y_{\overline{n}} - x_{\overline{n}} = (y_0 - x_0)/2^{\overline{n}} < \varepsilon$. Allora $[x_{\overline{n}}, y_{\overline{n}}] \subset [\overline{x} - \varepsilon, y_{\overline{n}}]$ in contraddizione col fatto che $[x_{\overline{n}}, y_{\overline{n}}]$ contiene punti di A. □

Possiamo enunciare il teorema ora dimostrato in una forma equivalente, ma forse più suggestiva. Sia $\{A, B\}$ una partizione di $\mathbb{R}$ (cioè A e B sono insiemi non vuoti

e disgiunti la cui unione è $\mathbb{R}$); essa si chiama **sezione** se: $\forall a \in A$ e $\forall b \in B$ risulta $a < b$.

Corollario 2.4 - *Per ogni sezione $\{A, B\}$ di $\mathbb{R}$ esiste un unico numero reale s (detto elemento separatore) tale che*

$$a \leq s \leq b \quad \forall a \in A \ , \quad \forall b \in B \ .$$

Per esercizio lo studente dimostrerà che s non è altro che $\sup A = \inf B$.

Nella presentazione assiomatica dei numeri reali, la proprietà $\mathcal{R}_4$, di solito enunciata nella forma del precedente corollario, prende il nome di *assioma di Dedekind* (o *di completezza* o *di continuità*). Pensando alla rappresentazione geometrica dei numeri sulla retta, osserviamo che l'assioma di Dedekind è l'analogo del *postulato di continuità della retta* in Euclide.

Direttamente dalla definizione di numero reale ricaviamo le seguenti proposizioni:

Proposizione 2.5 - $\mathbb{R}$ *possiede la* **proprietà di Archimede**: *per ogni coppia a, b di numeri reali positivi, esiste un intero naturale n tale che $na > b$.*

Infatti basterà prendere $n = [b/a] + 1$ (parte intera di $b/a + 1$) (*) .

Si dice che un sottoinsieme $T \subseteq \mathbb{R}$ è *denso* in $\mathbb{R}$ se: $\forall a, b \in \mathbb{R}$ con $a < b$, esiste $t \in T$ con $a < t < b$.

Si ha allora:

Proposizione 2.6 - *L'insieme dei razionali $\mathbb{Q}$ è denso in* $\mathbb{R}$.

2.5 Isomorfismo tra campi ordinati completi

Un insieme con le proprietà $\mathcal{R}_1$, $\mathcal{R}_2$, $\mathcal{R}_3$, $\mathcal{R}_4$ si dice **corpo commutativo (campo) ordinato completo.**

Ricordiamo che due campi ordinati C e C' sono *isomorfi* se esiste una corrispondenza biunivoca fra gli elementi $c \in C$ e gli elementi $c' \in C'$ tale che:

$$\left.\begin{array}{l} c \leftrightarrow c' \\ d \leftrightarrow d' \\ c < d \end{array}\right\} \Longrightarrow \left\{\begin{array}{l} c + d \leftrightarrow c' + d' \\ cd \leftrightarrow c'd' \\ c' < d' \end{array}\right.$$

cioè tale corrispondenza conserva la struttura algebrica e l'ordinamento.

Il seguente teorema di algebra, di cui non riportiamo la dimostrazione, mette ulteriormente in luce l'importanza della proprietà $\mathcal{R}_4$.

■ **Teorema 2.7** - *Due campi ordinati e completi sono isomorfi.*

(*) La parte intera $[a]$ di un numero reale a è il più grande intero che non supera a: $[\pi] = 3$, $[-\sqrt{3}] = -2$.

Esiste dunque, *sostanzialmente*, un *unico* campo ordinato completo, quello dei numeri reali. Questo teorema giustifica così anche la nostra presentazione dei numeri reali: fossimo partiti da allineamenti di cifre binarie o ternarie o altro (invece che decimali) avremmo ancora definito un campo ordinato completo, che però sarebbe risultato identificabile con quello sopra ottenuto. Così ogni altra costruzione dell'insieme dei numeri reali (utilizzando, per esempio, sezioni del campo razionale, oppure coppie di classi contigue di razionali, oppure successioni fondamentali di razionali) porta allo stesso "unico" risultato.

2.6 Potenza del continuo

Concludiamo la sezione mostrando che l'insieme $\mathbb{R}$ è strettamente "più grande" di $\mathbb{Q}$ nel senso che $\mathrm{card}(\mathbb{R}) > \mathrm{card}(\mathbb{Q}) = \mathrm{card}(\mathbb{N})$.

▪ **Teorema 2.8** - *L'insieme* $\mathbb{R}$ *non è numerabile.*

Per la dimostrazione facciamo uso della seguente proposizione, già interessante per sé stessa.

Lemma 2.9 - *Gli insiemi* $\mathbb{R}$ *e* $(0,1)$ *sono equipotenti.*

Dimostrazione - Osserviamo anzitutto che ogni intervallo $(a,b) \subset \mathbb{R}$ è equipotente a $(0,1)$: ad esempio l'applicazione $x \mapsto (b-a)x + a$ è biiettiva da $(0,1)$ su (a,b). Allora basterà mettere in corrispondenza biunivoca $\mathbb{R}$ con $(-1,1)$; un modo di realizzare una tale corrispondenza è illustrato nella figura sottostante:

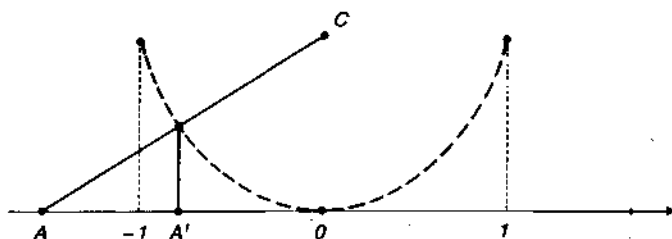



Fig. 2.1

Dimostrazione del teorema 2.8 - Grazie al Lemma 2.9 basterà provare che $(0,1)$ non è numerabile. Procediamo per assurdo. Sia $(0,1)$ numerabile, cioè esista una funzione $\varphi: \mathbb{N} \to (0,1)$ biiettiva che permette di ordinare i numeri di $(0,1)$ in una successione:

$$\begin{aligned} a_0 &= 0.\alpha_{00}\ \alpha_{01}\ \alpha_{02}\cdots \\ a_1 &= 0.\alpha_{10}\ \alpha_{11}\ \alpha_{12}\cdots \\ a_2 &= 0.\alpha_{20}\ \alpha_{21}\ \alpha_{22}\cdots \\ &\vdots \end{aligned}$$

Per arrivare ad una contraddizione basterà costruire un numero $b \in (0,1)$ diverso da tutti quelli della successione. Sia $b = 0.\beta_1\beta_2\beta_3\cdots$ ove le cifre decimali β_n sono così definite: $\beta_n = 7$ se $0 \le \alpha_{nn} \le 4$, $\beta_n = 3$ se $5 \le \alpha_{nn} \le 9$. Tale numero non può coincidere con alcuno degli a_n poiché $\alpha_{nn} \ne \beta_n$. □

Diremo che un insieme ha la cardinalità o *potenza del continuo* se risulta equipotente a $\mathbb{R}$.

Esercizi

1. Vero o falso:

a) La somma di due irrazionali è irrazionale.

b) La somma ed il prodotto tra un razionale ed un irrazionale è irrazionale.

2. Dimostrare che $\sqrt{3}$ e $\sqrt{5}$ sono irrazionali.

3. Un numero reale x si dice *algebrico* se è soluzione di un'equazione del tipo $a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 = 0$ dove ogni a_j $(j = 1,\ldots,n)$ è un intero relativo. Un numero reale *non algebrico* si dice *trascendente*. Sono algebrici tutti i numeri razionali e numeri irrazionali come $\sqrt{2}$, $\sqrt[3]{7}\ldots$; π è un classico esempio di irrazionale trascendente. Dimostrare che l'insieme dei numeri reali algebrici è numerabile.

4. Dimostrare che $\forall a, b, c, d, \in \mathbb{R}_+$

$$\min\left\{\frac{a}{c}, \frac{b}{d}\right\} \le \frac{a+b}{c+d} \le \max\left\{\frac{a}{c}, \frac{b}{d}\right\}.$$

5. Sia $[x]$ la parte intera di x. Dimostrare che $\forall x, y \in \mathbb{R}$ e $\forall n \in \mathbb{N}_+$:

a) $[x+y] \ge [x] + [y]$

b) $[x/n] = [[x]/n]$

c) $\sum\limits_{k=0}^{n-1} [x + k/n] = [nx]$.

6. Risolvere le seguenti disequazioni:

a) $|x-4| \ge x+2$

b) $\dfrac{(2x-1)(x+1)}{x} \ge 0$

c) $\dfrac{x}{|x|}(1-x) \le 1+x$.

7. Sia $A \subset \mathbb{Q}$ superiormente limitato. Allora $\sup A$ (vero o falso):

a) non è razionale,

b) è razionale solo se è massimo,

c) può essere razionale o irrazionale.

8. Quale dei seguenti insiemi ha come "minimo" 3?

a) $\{x \in \mathbb{R}: [x] = 3\}$

b) $\left\{x \in \mathbb{R}: 3 + \dfrac{2}{x+1}, x \in \mathbb{N}\right\}$

c) $\{x \in \mathbb{R}: x^3 \geq 27\}$.

9. Determinare estremo superiore e inferiore dei seguenti insiemi di $\mathbb{R}$, precisando ogni volta se si tratta di massimo e minimo

$$\left\{\frac{n-1}{n+1}\right\}_{n\in\mathbb{N}} , \quad \left\{x \in \mathbb{R}: x|x| < x^2\right\} \cup [-1,1) , \quad \{0\} , \quad \{x \in \mathbb{R}: 2 \leq x^2 < 4\} .$$

10. Sia $a \in \mathbb{R}_+$ e $a^{(n)}$ la sua troncata n-esima.
Dimostrare che la successione $a^{(n)} + 10^{-n}$ è non crescente ($a^{(n)} + 10^{-n}$ è l'approssimazione per eccesso di a all'n-esima cifra decimale).

11. Si deduca il principio di Archimede dalla proprietà $\mathcal{R}_4$ dei reali, senza far uso degli allineamenti decimali.
[Suggerimento: per assurdo, esistano due reali positivi a, b tali che, per ogni intero positivo n, risulti $na \leq b$; allora l'insieme $\{x \in \mathbb{R}: x = na, n \in \mathbb{N}\} \ldots$].

12. Si dimostri che $\mathbb{Q}$ è denso in $\mathbb{R}$ usando il principio di Archimede, senza far uso degli allineamenti decimali.

13.* Trovare un modo di mettere in corrispondenza biunivoca gli intervalli $(0,1)$ e $[0,1]$.

Appendice A: Dimostrazione delle proprietà di campo ordinato dei numeri reali

È sufficiente limitarsi a numeri *reali positivi*. Siano dunque $a, b, c \in \mathbb{R}_+$.
Cominciamo dalle proprietà della somma (proprietà $\mathcal{R}_1$).

1. $a + b = b + a$ (commutatività)
2. $(a + b) + c = a + (b + c)$ (associatività)
3. $a + 0 = 0 + a = a$ (esistenza dell'elemento neutro)
4. $a + (-a) = (-a) + a = 0$ (esistenza dell'opposto) .

Le 3. e 4. seguono direttamente dalla definizione.
Per la 1. è sufficiente osservare che $a^{(n)} + b^{(n)} = b^{(n)} + a^{(n)}$ (commutatività per i numeri razionali).
Per la 2. osserviamo che:

$$a^{(n)} + b^{(n)} + c^{(n)} \leq (a+b)^{(n)} + c^{(n)} \leq a^{(n)} + b^{(n)} + c^{(n)} + 10^{-n}$$

e

$$a^{(n)} + b^{(n)} + c^{(n)} \leq a^{(n)} + (b+c)^{(n)} \leq a^{(n)} + b^{(n)} + c^{(n)} + 10^{-n}$$

e quindi, posto

$$x^{(n)} = (a+b)^{(n)} + c^{(n)} \quad \text{e} \quad y^{(n)} = a^{(n)} + (b+c)^{(n)}$$

si ha che $x^{(n)} \rightrightarrows (a+b)+c$ mentre $y^{(n)} \rightrightarrows a+(b+c)$; inoltre:

$$-10^{-n} \leq x^{(n)} - y^{(n)} \leq 10^{-n} \,. \tag{A.1}$$

La conclusione si ottiene dal seguente lemma (con $M = 1$).

Lemma A.1 - *Siano $x^{(n)} \rightrightarrows x \geq 0$, e $y^{(n)} \rightrightarrows y \geq 0$. Supponiamo che esista un numero reale positivo M tale che per ogni $n \geq 1$ si abbia*

$$-M \cdot 10^{-n} < x^{(n)} - y^{(n)} < M \cdot 10^{-n} \,. \tag{A.2}$$

Allora $x = y$.

Dimostrazione - Siano

$$x = \alpha_0 \,.\, \alpha_1 \alpha_2 \alpha 3 \cdots \quad , \quad y = \beta_0 \,.\, \beta_1 \beta_2 \beta_3 \cdots \,.$$

Supponiamo $x > y$ e poniamo $r = x - y$, $r^{(n)} = x^{(n)} - (y^{(n)} + 10^{-n})$. Dalla definizione di differenza abbiamo che $r^{(n)} \rightrightarrows r$. Sia k tale che $2 \cdot 10^{-k} < r$.

Per definizione di successione stabilizzata esisterà n_k tale che, se $n > n_k$, le prime k cifre (dopo il punto) di $r^{(n)}$ sono stabilizzate; dunque se $n > n_k$ si ha $r - r^{(n)} < 10^{-k}$ ovvero

$$x^{(n)} - y^{(n)} > r - 2 \cdot 10^{-k} \,. \tag{A.3}$$

Scegliamo ora $n_0 > n_k$ in modo che $M \cdot 10^{-n_0} < r - 2 \cdot 10^{-k}$. Le (A.2) ed (A.3) per $n = n_0$ sono in contrasto tra loro. Dunque non può essere $x > y$.

Analogamente si ragiona per dimostrare che non può essere $x < y$. □

Passiamo ora alle proprietà del prodotto (Proprietà $\mathcal{R}_2$).

1. $a \cdot b = b \cdot a$ (commutatività)
2. $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ (associatività)
3. $a \cdot 1 = 1 \cdot a = a$ (esistenza dell'elemento neutro)
4. $a \cdot \left(\frac{1}{a}\right) = \left(\frac{1}{a}\right) \cdot a = 1 \quad (a \neq 0)$ (esistenza dell'inverso) .

La 1. segue da $a^{(n)} \cdot b^{(n)} = b^{(n)} \cdot a^{(n)}$ (commutatività del prodotto tra razionali).
La 3. segue dal fatto che $1^{(n)} = 1$.

Per la 2. osserviamo che, posto $M = \max\{c\ (a+b+2), a\ (b+c+2)\}$ si ha:

$$a^{(n)} \cdot b^{(n)} \cdot c^{(n)} - M \cdot 10^{-n} \leq ((a\ b)^{(n)} \cdot c^{(n)})^{(n)} \leq a^{(n)} \cdot b^{(n)} \cdot c^{(n)} + \\ + M \cdot 10^{-n} \ ,$$

$$a^{(n)} \cdot b^{(n)} \cdot c^{(n)} - M \cdot 10^{-n} \leq (a^{(n)} \cdot (b\ c)^{(n)})^{(n)} \leq a^{(n)} \cdot b^{(n)} \cdot c^{(n)} + \\ + M \cdot 10^{-n} \ .$$

Dal lemma A.1 poiché

$$((a\ b)^{(n)} \cdot c^{(n)})^{(n)} \rightrightarrows (a\ b)\ c \quad \text{e} \quad (a^{(n)} \cdot (b\ c)^{(n)})^{(n)} \rightrightarrows a(b\ c)$$

si ottiene la 2.

Infine, poniamo

$$q^{(n)} = \left(\left(\frac{1}{a} \right)^{(n)} \cdot a^{(n)} \right)^{(n)} .$$

Poiché

$$\frac{1}{a^{(n)} + 10^{-n}} \rightrightarrows \frac{1}{a} \ ,$$

per ogni n esiste s_0 tale che se $s \geq s_0$

$$\left(\frac{1}{a} \right)^{(n)} = \left(\frac{1}{a^{(s)} + 10^{-s}} \right)^{(n)} < \frac{1}{a^{(s)} + 10^{-s}} < \frac{1}{a^{(s)}} \ . \tag{A.4}$$

Scegliamo ora anche s maggiore di n; si ricava così:

$$q^{(n)} \underset{\substack{\uparrow \\ \text{dalla (A.4)}}}{<} \left(\frac{1}{a^{(s)}} \cdot a^{(n)} \right)^{(n)} \underset{\substack{\uparrow \\ \text{poiché } s > n}}{<} \left(\frac{1}{a^{(n)}} \cdot a^{(n)} \right)^{(n)} = 1 \tag{A.5}$$

D'altra parte, con la stessa scelta di s si ha:

$$\left(\frac{1}{a} \right)^{(n)} = \left(\frac{1}{a^{(s)} + 10^{-s}} \right)^{(n)} > \frac{1}{a^{(s)} + 10^{-s}} - 10^{-n} > \\ > \frac{1}{a^{(s)} + 10^{-n}} - 10^{-n} \ .$$

Essendo $a^{(s)} < a^{(n)} + 10^{-n}$ si ha:

$$\left(\frac{1}{a} \right)^{(n)} \cdot a^{(n)} > (a^{(s)} - 10^{-n}) \cdot \left[\frac{1}{a^{(s)} + 10^{-n}} - 10^{-n} \right] > \\ > 1 - \left[\frac{2}{a^{(0)}} + a \right] \cdot 10^{-n} \ . \tag{A.6}$$

Dalle (A.5) e (A.6) si ottiene per ogni $n \geq 1$:

$$1 - \left[\frac{2}{a^{(0)}} + a\right] \cdot 10^{-n} < q^{(n)} < 1 \,,$$

ovvero

$$- M \cdot 10^{-n} < q^{(n)} - 1 < 0 \qquad \left(M = \left[\frac{2}{a^{(0)}} + a\right]\right) .$$

Dal lemma A.1 si ottiene la 4.

La proprietà *distributiva* $(a+b)c = ac+bc$ si dimostra in modo analogo; lasciamo i dettagli al lettore.

Veniamo ora alle proprietà di compatibilità $\mathcal{R}_3$:

1. $a < b \Longrightarrow a + c < b + c \qquad \forall\, a, b, c$
2. $a < b \Longrightarrow ac < bc \qquad \forall\, a, b, c$ con $c > 0$.

Dimostriamo la 1.
Siano r_1 ed r_2 due numeri razionali tali che

$$a < r_1 < r_2 < b \,.$$

Questi numeri si possono trovare poiché $\mathbb{Q}$ è *denso* in $\mathbb{R}$.
Per i razionali 1. è vera; dunque $r_1 + c^{(n)} < r_2 + c^{(n)}$, per ogni n.
Per n grande si ha:

$$a^{(n)} < r_1 < r_2 < b^{(n)}$$

e

$$a^{(n)} + c^{(n)} < r_1 + c^{(n)} < r_2 + c^{(n)} < b^{(n)} + c^{(n)} \,.$$

Poiché $r_1 + c^{(n)} \rightrightarrows r_1 + c$ e $r_2 + c^{(n)} \rightrightarrows r_2 + c$ otteniamo subito

$$r_1 + c < r_2 + c \,.$$

Da $a^{(n)} + c^{(n)} \rightrightarrows a + c$ e $a^{(n)} + c^{(n)} < r_1 + c^{(n)} < r_1 + c$ otteniamo $a + b <$ $< r_1 + c$.
Analogamente $r_2 + c < b + c$.
La 1. segue ora facilmente.
Per la 2. si opera in maniera analoga, dunque tralasciamo i dettagli.

Appendice B: I numeri-macchina. Errori

Utilizzando una normale calcolatrice tascabile proviamo ad eseguire il calcolo di

$$(\sqrt[3]{2} + 100'000 - 100'000)^3 \ .$$

La risposta esatta è 2 ma, molto probabilmente, la vostra calcolatrice fornirà un valore diverso; ciò è dovuto al fatto che essa (ma anche il più potente dei computers!) si serve di un numero finito di cifre, troncando o arrotondando le cifre decimali di ogni numero. Pertanto i numeri reali (razionali e irrazionali) che sono allineamenti illimitati di cifre, non possono essere rappresentati su un calcolatore. Questo opera su un insieme *finito* di numeri, detti *numeri-macchina*. Grosso modo, un tal numero viene scritto nella forma αq^h, dove α si chiama mantissa e q è la base del sistema di numerazione usato dalla macchina. Il numero delle cifre disponibili per la mantissa e per l'esponente della base usata per la rappresentazione dei numeri varia da macchina a macchina.

Più precisamente, nella rappresentazione dei numeri detta "*in virgola mobile*" *(floating point representation)*, la macchina opera con numeri della forma

$$\pm\, q^h \cdot \sum_{k=1}^{t} \alpha_k q^{-k} = \pm q^h \cdot \underbrace{0.\alpha_1\alpha_2\cdots\alpha_t}_{\alpha} \ .$$

L'esponente h può avere segno qualunque ed è soggetto ad una limitazione del tipo $|h| \leq h_0$. Tale rappresentazione è unica se si richiede che $\alpha_1 \neq 0$ ovvero che $\alpha \geq q^{-1}$; in questo caso il numero si dice scritto in *forma normalizzata.*

Ad esempio i numeri

$$321.33 \cdot 10^3 \quad \text{e} \quad 0.0314 \cdot 10^{-2}$$

sono scritti in virgola mobile con $q = 10$; in *forma normalizzata* gli stessi numeri si scrivono rispettivamente,

$$0.32133 \cdot 10^6 \quad \text{e} \quad 0.314 \cdot 10^{-3} \ .$$

A titolo d'esempio eseguiamo la sottrazione tra

$$a = 0.1002 \cdot 10^{-3} \quad \text{e} \quad b = 0.9995 \cdot 10^{-4}$$

con una macchina che opera con $q = 10$ e $t = 4$

Per effettuare $a - b$, la macchina sceglie il massimo tra gli esponenti -3 e -4, quindi -3 e, scritto $b = 0.09995 \cdot 10^{-3}$ arrotonda la mantissa alla 4^a cifra decimale ottenendo $b' = 0.1000 \cdot 10^{-3}$.

Si ha:

$$a - b = 0.25 \cdot 10^{-6} \ ,$$

mentre la macchina esegue

$$a - b' = 0.2 \cdot 10^{-6} \ .$$

Il risultato finale del calcolo sarà perciò una approssimazione del valore esatto.

Se x è il valore esatto, e $\overline{x}$ una sua approssimazione, si definiscono gli errori assoluto e relativo nel modo seguente:

Errore assoluto: $\Delta = |x - \overline{x}|$

Errore relativo: $\delta = \Delta/|\overline{x}| = \dfrac{|x - \overline{x}|}{|\overline{x}|}$.

Ad esempio, la scrittura

$$x = 0.128 \pm 0.004 = 0.128 \pm 4 \cdot 10^{-3}$$

con la quale vogliamo indicare il risultato di un calcolo, indica che:

$$0.128 - 0.004 \leq x \leq 0.128 + 0.004$$

cioè

$$0.124 \leq x \leq 0.132 \ .$$

In tal caso risulta $\Delta = 4 \cdot 10^{-3}$, $\delta = 31.25 \cdot 10^{-3}$.
In generale si scriverà

$$x = \overline{x} \pm \Delta$$
$$x = \overline{x}(1 \pm \delta) \ .$$

Definiamo ora l'errore di *troncamento* e di *arrotondamento*.
Il numero π è noto oggi con 15 milioni di cifre decimali

$$\pi = 3.14159265358979323846 26 \cdots$$

I numeri

3.1 3.14 3.141 3.1415 ...

ottenuti considerando nell'allineamento di π solo la prima, le prime 2, le prime 3, ... cifre decimali si dicono valori di π *troncati* rispettivamente alla prima, seconda, terza etc. cifra decimale. Considerando un valore troncato alla n-esima cifra decimale si commette un errore non superiore a 10^{-n}.
Analogamente i numeri

3.1 3.14 3.142 3.1416 ...

ottenuti troncando l'allineamento come si è fatto in precedenza, ma aumentando di una unità l'ultima cifra considerata se quella successiva è ≥ 5, si dicono *arrotondati* rispettivamente alla prima, seconda, etc. cifra decimale. Arrotondando all'n-esima cifra decimale si commette un errore non superiore a $0.5 \cdot 10^{-n}$.

Il fatto di operare su numeri abbreviati (cioè troncati o arrotondati) comporta che le operazioni algebriche svolte sull'insieme dei numeri-macchina non posseggono più tutte le proprietà strutturali che posseggono quando esse sono definite sul campo dei numeri reali. Ad esempio, supponiamo di operare con numeri 3-ridotti (cioè arrotondati alla 3^a cifra decimale). Siano

$$a = 0.534 \quad b = 0.704 \quad c = 0.025 .$$

Calcoliamo

$$(ab)c = 0.009$$
$$a(bc) = 0.01 .$$

Dunque la proprietà associativa cade in difetto!

Uno dei problemi più importanti dell'Analisi Numerica è il controllo degli errori nei procedimenti di calcolo. Come si propagano gli errori iniziali nel corso dei calcoli? Come controllare gli errori generati da troncamenti o arrotondamenti? Se è noto l'errore nei dati iniziali, con quale precisione è noto il risultato finale? E, viceversa, se si vuole un risultato finale con una data precisione, con quanta precisione devono essere noti i dati iniziali?

Esaminiamo, a titolo d'esempio, le operazioni elementari. Siano a e b numeri *positivi* e $\overline{a}$ e $\overline{b}$ una loro approssimazione:

$$a = \overline{a} \pm \Delta_a \ , \ a = \overline{a}(1 \pm \delta_a)$$
$$b = \overline{b} \pm \Delta_b \ , \ b = \overline{b}(1 \pm \delta_b) .$$

Somma: risulta immediatamente $a + b = \overline{a} + \overline{b} \pm (\Delta_a + \Delta_b)$ cioè

$$\Delta_{a+b} \leq \Delta_a + \Delta_b . \qquad (B.1)$$

In altre parole, gli errori assoluti si sommano.

Veniamo all'errore relativo; si ha

$$\frac{|a + b - (\overline{a} + \overline{b})|}{\overline{a} + \overline{b}} = \frac{|(a - \overline{a}) + (b - \overline{b})|}{\overline{a} + \overline{b}} \leq \frac{|\overline{a} - a|}{\overline{a} + \overline{b}} + \frac{|\overline{b} - b|}{\overline{a} + \overline{b}} = (^*)$$
$$= \frac{\overline{a}}{\overline{a} + \overline{b}} \frac{|a - \overline{a}|}{\overline{a}} + \frac{\overline{b}}{\overline{a} + \overline{b}} \frac{|b - \overline{b}|}{\overline{b}} .$$

(*) Utilizziamo qui la "disuguaglianza triangolare" che sarà dimostrata in 3.4: se x, y sono numeri reali, $|x + y| \leq |x| + |y|$.

Risulta dunque

$$\delta_{a+b} \leq \frac{\overline{a}}{\overline{a}+\overline{b}}\,\delta_a + \frac{\overline{b}}{\overline{a}+\overline{b}}\,\delta_b\,. \tag{B.2}$$

Naturalmente, essendo i numeri

$$\frac{\overline{a}}{\overline{a}+\overline{b}}\,,\ \frac{\overline{b}}{\overline{a}+\overline{b}} \tag{B.3}$$

minori di 1, risulta ancora

$$\delta_{a+b} \leq \delta_a + \delta_b\,. \tag{B.4}$$

La formula (B.2) mette in evidenza la sensibilità dell'errore finale rispetto ai numeri (B.3), che sono detti *indici di condizionamento.*

***Esempio* B.1** - Per calcolare $\pi + \sqrt{3}$ con almeno 6 cifre decimali esatte dopo il punto, con quale precisione dovremo calcolare π e $\sqrt{3}$?
Utilizziamo valori arrotondati: se $\pi = a \pm 0.5 \cdot 10^{-n}$ e $\sqrt{3} = b \pm 0.5 \cdot 10^{-n}$, allora $\pi + \sqrt{3} = a + b \pm 10^{-n}$. Se vogliamo almeno 6 cifre decimali esatte possiamo prendere $n = 7$. Quindi $\pi = 3.1415927 \pm 0.5 \cdot 10^{-7}$, $\sqrt{3} = 1.7320508 \pm 0.5 \cdot 10^{-7}$ e $\pi + \sqrt{3} = 4.8736435 \pm 10^{-7}$, da cui

$$4.8736434 < \pi + \sqrt{3} < 4.8736436$$

Le 6 cifre decimali 873643 sono dunque esatte.

Differenza: valgono considerazioni analoghe a quelle svolte per la somma. In particolare si trova

$$\begin{aligned} &\Delta_{a-b} \leq \Delta_a + \Delta_b \\ &\delta_{a-b} \leq \frac{\overline{a}}{|\overline{a}-\overline{b}|}\,\delta_a + \frac{\overline{b}}{|\overline{a}-\overline{b}|}\,\delta_b\,. \end{aligned} \tag{B.5}$$

Gli indici di condizionamento sono ora i numeri

$$\frac{\overline{a}}{|\overline{a}-\overline{b}|}\,,\ \frac{\overline{b}}{|\overline{a}-\overline{b}|}$$

che mettono in evidenza la possibilità di un'amplificazione degli errori relativi quando si sottraggono quantità molto vicine tra loro.

Prodotto: si trova subito che

$$\Delta_{ab} \leq \overline{a}\Delta_b + \overline{b}\Delta_a + \Delta_a\Delta_b\,. \tag{B.6}$$

Spesso $\Delta_a\Delta_b \ll 1$ ($\ll$ significa: molto minore) e la (B.6) si approssima trascurando $\Delta_a\Delta_b$:

$$\Delta_{ab} \leq \overline{\Delta}_{ab} \simeq \overline{a}\Delta_b + \overline{b}\Delta_a \ . \tag{B.7}$$

Le formule (B.6) e (B.7) mostrano già che il prodotto si comporta in modo molto peggiore della somma per quanto riguarda la propagazione degli errori non appena intervengono fattori abbastanza grandi.

***Esempio* B.2** - Si voglia calcolare il numero $\pi\sqrt{7}$ a partire dai valori di π e $\sqrt{7}$ arrotondati alla 4^a cifra decimale: $\pi = 3.1416 \pm 0.5\cdot 10^{-4}$ e $\sqrt{7} = 2.6458 \pm 0.5 \cdot 10^{-4}$. Avremo allora: $\pi\sqrt{7} = 8.31204528 \pm \Delta_{ab}$ con $\Delta_{ab} \simeq 0.00028937$ (si è trascurato $\Delta_a\Delta_b \simeq 0.25 \cdot 10^{-8} \ll 1$). Di conseguenza potremo scrivere

$$8.31175592 \leq \pi\sqrt{7} \leq 8.31233464$$

e cioè $\pi\sqrt{7} = 8.31\ldots$ con la terza cifra decimale incerta tra 1 e 2!

Una formula analoga alla (B.6) si trova per l'errore relativo:

$$\delta_{ab} \leq \delta_a + \delta_b + \delta_a\delta_b \ . \tag{B.8}$$

Quoziente: per l'errore assoluto si ha:

$$\Delta_{\frac{a}{b}} = \left|\frac{a}{b} - \frac{\overline{a}}{\overline{b}}\right| = \left|\frac{\overline{a} \pm \Delta_a}{\overline{b} \pm \Delta_b} - \frac{\overline{a}}{\overline{b}}\right| = \left|\frac{\pm\overline{b}\Delta_a \pm \overline{a}\Delta_b}{\overline{b}(\overline{b} \pm \Delta_b)}\right|$$

da cui la formula

$$\Delta_{\frac{a}{b}} \leq \frac{\overline{a}\Delta_b + \overline{b}\Delta_a}{\overline{b}(\overline{b} - \Delta_b)} \tag{B.9}$$

formula che si semplifica approssimando il denominatore con $\overline{b}^2$; si ha così

$$\Delta_{\frac{a}{b}} \leq \frac{\overline{a}\Delta_b + \overline{b}\Delta_a}{\overline{b}^2} \ . \tag{B.10}$$

***Esempio* B.3** - Sia $e = 2.71828 \pm 0.5 \cdot 10^{-5}$ e $\sqrt{11} = 3.31662 \pm 0.5 \cdot 10^{-5}$. Si voglia calcolare $e/\sqrt{11}$; utilizzando la (B.9) otteniamo:

$$\Delta_{\frac{e}{\sqrt{11}}} \leq \frac{(3.31662 + 2.71828)0.5 \cdot 10^{-5}}{3.31662 \cdot 3.316615} \leq 2.8 \cdot 10^{-6} \ .$$

E quindi

$$\frac{e}{\sqrt{11}} = \frac{2.71828}{3.31662} \pm 2.8 \cdot 10^{-6} = 0.8195934415 \pm 2.8 \cdot 10^{-6} \ .$$

In conclusione abbiamo

$$0.8195906 \leq \frac{e}{\sqrt{11}} \leq 0.8195962$$

che dà il valore di $e/\sqrt{11}$ con cinque decimali esatti!

Una formula analoga alla (B.9) vale per l'errore relativo:

$$\delta_{a/b} = \frac{\Delta_{a/b}}{\overline{a}/\overline{b}} \leq \frac{\overline{b}}{\overline{a}} \frac{\overline{a}\Delta_b + \overline{b}\Delta_a}{\overline{b}(\overline{b} - \Delta_b)} = \frac{\delta_a + \delta_b}{1 - \delta_b} \tag{B.11}$$

oppure, trascurando δ_b rispetto a 1,

$$\delta_{a/b} \leq \delta_a + \delta_b \,. \tag{B.12}$$

L'analisi dell'errore in un problema computazionale e la stima dei suoi indici di condizionamento (di cui sopra abbiamo visto qualche esempio tra i più semplici) portano a distinguere tra problemi "*stabili*" o "*ben condizionati*" (per i quali "piccoli" errori nei dati producono "piccoli" errori nei risultati) e problemi "*instabili*" o "*mal condizionati*" (per i quali "piccoli" errori nei dati producono "grandi" errori nei risultati).

Sarebbe perciò auspicabile (anche se generalmente difficile e noioso) che una tale analisi fosse sviluppata prima di inziare l'effettivo procedimento di calcolo.

Queste poche note vogliono solo mettere in guardia l'allievo dall'uso, in maniera acritica, del più potente strumento di calcolo che la tecnica oggi ci mette a disposizione.

3. RADICALI - POTENZE - LOGARITMI

3.1 Radici n-esime aritmetiche

In conseguenza del teorema di completezza possiamo eseguire nel campo reale, operazioni che sono solo occasionalmente possibili nel campo razionale.

Consideriamo l'equazione $x^2 = y$; essa non ha alcuna soluzione se $y < 0$ (in virtù della struttura di campo ordinato esistente in $\mathbb{R}$), e ha evidentemente la sola soluzione nulla se $y = 0$; possiede invece, come vedremo, esattamente due soluzioni x_1 e x_2 se $y > 0$, l'una positiva e l'altra negativa. È evidente che ci basterà mostrare l'esistenza di una soluzione positiva x_1 e allora seguirà subito necessariamente $x_2 = -x_1$.

Invece, l'equazione $x^3 = y$ ha sempre, per ogni y, una sola soluzione x_1; risulta $x_1 = 0$ se $y = 0$ (evidente!), $x_1 > 0$ se $y > 0$ (come ora vedremo) e $x_1 < 0$ se $y < 0$. Quest'ultima affermazione segue da quella relativa al caso $y > 0$, perché, se abbiamo provato che $x_1 > 0$ è l'unica soluzione di $x^3 = y$ (con $y > 0$) segue subito che $-x_1$ è l'unica soluzione di $x^3 = -y$.

Lo studente non avrà difficoltà poi a ripetere la discussione per l'equazione $x^n = y$, separando il caso n *pari* da quello n *dispari*. La ricerca di queste soluzioni corrisponde all'operazione di estrazione di radice n-esima.

■ **Teorema 3.1** - *Sia* $y \in \mathbb{R}$, $y > 0$ *e* n *intero* ≥ 1. *Esiste un unico numero reale positivo* x *tale che* $x^n = y$.

Tale numero si chiama **radice n-esima aritmetica** di y e si indica con uno dei simboli $\sqrt[n]{y}$ oppure $y^{1/n}$.

Dimostrazione (per $n = 2$) - L'idea è di costruire una successione stabilizzata di razionali, che individui un numero reale x con le proprietà richieste.

L'unicità di un tale x segue subito dal fatto che, se ne esistessero due, $0 < x_1 < x_2$ sarebbe anche $0 < x_1^2 < x_2^2$: assurdo, poiché $x_1^2 = x_2^2 = y$.

Cominciamo col porre $a_0 = p$, dove p è il massimo intero tale che $p^2 \leq y$. Evidentemente $p \geq 0$. Se $p^2 = y$ allora $x = p$; se $p^2 < y$, poniamo $a_1 = p.\alpha_1$ dove α_1 è la massima tra le cifre decimali c_s tali che $(p.c_s)^2 \leq y$. Evidentemente $a_1 \geq p$. Se $a_1^2 = y$ allora $x = a_1$, altrimenti si prosegue ponendo $a_2 = p.\alpha_1\alpha_2$ dove α_2 è la massima tra le cifre decimali c_s tali che $(p.\alpha_1 c_s)^2 \leq y$. Evidentemente $a_2 \geq a_1$. Proseguendo in questo modo otteniamo una successione di razionali $\{a_k\}$:

$$a_k = p.\alpha_1\alpha_2 \cdots \alpha_k$$

tale che: (1) per un certo $\overline{k}$, $(a_{\overline{k}})^2 = y$, oppure: (2) per ogni k, $a_k^2 < y$ e α_k è la massima tra le cifre decimali c_s tali che $(p.\alpha_1\alpha_2 \cdots \alpha_{k-1}c_s)^2 < y$.

Nel caso (1) si pone $x = a_{\overline{k}}$ e il teorema è dimostrato. Nel caso (2) osserviamo che $\{a_k\}$ è non decrescente e limitata superiormente da $p + 1$. Per il Lemma 2.1, resta individuato un numero x tale che $a_k \rightrightarrows x$. Mostriamo che $x^2 = y$.

Poiché $x^{(k)} = a_k$ (cioè troncando x alla k-esima cifra decimale si ottiene a_k) si ha

$$(a_k \cdot a_k)^{(k)} \rightrightarrows x^2$$

ed essendo $(a_k \cdot a_k)^{(k)} \leq a_k \cdot a_k < y$, sempre dal citato Lemma si ricava $x^2 \leq y$.

Supponiamo ora per assurdo, che $x^2 < y$. Poniamo $y - x^2 = r > 0$, e scegliamo k_0 tale che risulti $r > (2x + 1)/10^{k_0}$. Allora avremmo

$$\begin{aligned}(a_{k_0} + 10^{-k_0})^2 &\leq (x + 10^{-k_0})^2 = x^2 + 2x\ 10^{-k_0} + 10^{-2k_0} < \\ &< x^2 + \frac{2x+1}{10^{k_0}} < x^2 + r = y\ .\end{aligned}$$

Dunque la cifra k_0-esima di a_{k_0} *non* è la massima tra le cifre decimali c_s tali che $(p.\alpha_1\alpha_2 \cdots \alpha_{k_0-1}c_s)^2 \leq y$. Avendo raggiunto una contraddizione, deduciamo che $x^2 = y$. Analogamente si prova il teorema per $n > 2$. □

Dalla dimostrazione è chiaro che $x = \sup\{a_k\}$. È quindi la proprietà di completezza che entra essenzialmente in gioco nel teorema 3.1.

Esempio **3.1** - La dimostrazione del teorema 3.1 indica come costruire la radice n–esima di un numero positivo. Per esempio per costruire $\sqrt{2}$ si procede così: essendo $1^2 < 2$ e $2^2 > 2$ si ha $p = 1$; essendo $(1.4)^2 = 1.96 < 2$ e $(1.5)^2 = 2.25 > 2$ si ha $a_1 = 1.4$; essendo $(1.41)^2 = 1.9881 < 2$ e $(1.42)^2 = 2.0164 > 2$ si ha $a_2 = 1.41$ e così di seguito.

3.2 Potenze con esponente reale

L'operazione di elevamento a potenza: a^b (si legge: "a elevato a b"; a è la base, b l'esponente) si definisce senza difficoltà se l'esponente è un intero ($a^3 := a \cdot a \cdot a$, $a^{-4} := 1/a^4$; $a^0 := 1$ per ogni $a \neq 0$) e si definisce per *ogni* esponente razionale solo se la base è positiva, utilizzando il teorema 3.1. Basterà limitarsi al caso $a \geq 1$, poiché, se $a < 1$, porremo:

$$a^b := \left(\frac{1}{a}\right)^{-b} . \tag{3.1}$$

Sia dunque $a \geq 1$, $r = m/n$, $m \in \mathbb{Z}$, $n \in \mathbb{N}_+$, m e n *primi fra loro*; poniamo

$$a^r := (a^m)^{\frac{1}{n}} = (\sqrt[n]{a})^m . \tag{3.2}$$

Dalle definizioni (3.1) e (3.2) discendono le note regole sugli esponenti, che si dimostrano senza difficoltà; siano a, b *reali positivi*, r, s *razionali*:

E_0 $\quad a^r > 0 \ \forall r; \ a^r \lesseqgtr 1$ se $a \lesseqgtr 1$, $r > 0$

E_1 $\quad a^{r+s} = a^r \cdot a^s$

E_2 $\quad (ab)^r = a^r \cdot b^r$

E_3 $\quad (a^r)^s = a^{r \cdot s}$

E_4 $\quad r < s \Longrightarrow \begin{array}{l} a^r < a^s \text{ se } a > 1 \\ a^r > a^s \text{ se } a < 1 \end{array}$

Notiamo che $1^r = 1$, $\forall r \in \mathbb{Q}$. Vale anche la seguente proprietà (immediata conseguenza di E_0) che permette di confrontare potenze con basi diverse.

E_5 $\quad 0 < a \leq b \Longrightarrow a^r \leq b^r$, $r > 0$.

Vogliamo ora estendere la definizione al caso di esponente reale. Usiamo la solita tecnica delle successioni stabilizzate.

Sia $a \geq 1$ e $b \in \mathbb{R}_+$. La successione di reali positivi $a^{b^{(n)}}$, dove $b^{(n)}$ è la troncata n-esima di b, è ben definita, non decrescente e limitata da $a^{[b]+1}$. Pertanto è stabilizzata e possiamo *definire* a^b *come il numero reale positivo individuato da* $a^{b^{(n)}}$ *cioè tale che:*

$$a^{b^{(n)}} \rightrightarrows a^b . \tag{3.3}$$

Si usa ancora la (3.1) per estendere la definizione alle basi *positive* < 1. Le proprietà $E_0 \dots E_5$ continuano a valere (come subito si accerta passando agli esponenti troncati).

3.3 Logaritmi

Consideriamo l'equazione

$$a^x = y \tag{3.4}$$

con $a > 0$. Anzitutto, se $a = 1$, essa è soddisfatta solo se $y = 1$ (e in tal caso ogni numero reale è soluzione). Sia dunque $a \neq 1$. Per la proprietà E_0 in 3.2 (estesa agli esponenti reali) essa non ha alcuna soluzione se $y \leq 0$: mostreremo che *ha una sola soluzione per ogni* $y > 0$. Questa soluzione prende il nome di **logaritmo in base** a **di** y e si indica col simbolo $\log_a y$.

Potremo limitarci a provare l'affermazione per $a > 1$ e $y \geq 1$; infatti i casi mancanti possono essere completati osservando che la (3.4) può scriversi anche nella forma:

$$\left(\frac{1}{a}\right)^{-x} = y \quad \text{oppure} \quad a^{-x} = \frac{1}{y} \ ;$$

in questo modo ci si riporta sempre al caso $a > 1$, $y \geq 1$.

■ **Teorema 3.2** - *Sia* $a > 1$, $y \geq 1$. *Esiste un unico numero reale* $x > 0$ *tale che* $a^x = y$.

Dimostrazione - L'unicità segue subito dalla proprietà E_4 di 3.2. Per mostrare l'esistenza, si costruisce, con la tecnica usata nella dimostrazione del teorema 3.1, una successione di razionali $\{r_k\}$ tale che:

(1) per un certo $\overline{k}$ risulta $a^{r_{\overline{k}}} = y$
oppure
(2) $a^{r_k} < y$ per ogni k e $a^{r_k + 10^{-k}} > y$.

Nel primo caso si pone $x = r_{\overline{k}}$ e il teorema è dimostrato.

Nel secondo caso si vede che $\{r_k\}$ è stabilizzata e perciò individua un unico numero reale x, con $r_k \leq x$ e $r_k = x^{(k)}$ per ogni k. Poiché $a^{r_k} \rightrightarrows a^x$ si ricava $a^x \leq y$. Sia, per assurdo, $a^x < y$. Poniamo $y - a^x = \alpha$ e scegliamo k tale che $a^x(a-1)/k < \alpha$. Otteniamo, essendo $10^{-k} < 1/k$ e usando la E_4:

$$a^{r_k + 10^{-k}} < a^x \, a^{10^{-k}} < a^x \, a^{\frac{1}{k}} \ .$$

Usiamo ora la seguente disuguaglianza

$$a^{\frac{1}{k}} \leq \left(1 + \frac{a-1}{k}\right) \quad (a > 1, k \in \mathbb{N}_+) \tag{3.5}$$

(per comodità, la dimostrazione di (3.5) è riportata nel par. 3.4 formula (3.12')). Otteniamo dunque:

$$a^{r_k+10^{-k}} < a^x\left(1+\frac{a-1}{k}\right) = a^x + a^x\frac{a-1}{k} < a^x + \alpha = y$$

che contraddice il modo con cui è stato costruito r_k. Perciò si conclude $a^x = y$. □

Dalle proprietà E_1, E_2, E_3, e direttamente dalla definizione si ricavano poi le seguenti regole, *valide per* x, y, a *positivi*, $a \neq 1$:

L_1 $\log_a xy = \log_a x + \log_a y$

L_2 $\log_a \dfrac{x}{y} = \log_a x - \log_a y$

L_3 $\log_a x^\alpha = \alpha \log_a x \quad \forall\alpha \in \mathbb{R}$

L_4 $\log_a x = \dfrac{1}{\log_x a} = -\log_{\frac{1}{a}} x \quad (x \neq 1)$.

Le operazioni definite in questa sezione consentono di introdurre tre importanti classi di funzioni (reali di variabile reale): le *potenze* (con esponente reale $\alpha \neq 0$): $x \mapsto x^\alpha$, gli *esponenziali* (con base positiva $a \neq 1$): $x \mapsto a^x$ e i *logaritmi* (con base positiva $a \neq 1$): $x \to \log_a x$, le cui proprietà saranno studiate più avanti (cfr. Cap. 5), ma i cui diagrammi devono già essere familiari allo studente.

Qui notiamo soltanto che permangono insoddisfacenti lacune nelle definizioni ora date. Per esempio, l'operazione di elevamento a potenza, con esponente reale, è definita solo se la base è > 0. Se la base è < 0 la definizione può estendersi facilmente se l'esponente è intero oppure anche razionale m/n (m e n primi fra loro) purché n sia dispari; diversamente non è possibile definire la potenza (in modo che si conservino le usuali regole di calcolo). Questa e altre limitazioni saranno completamente rimosse (ma pagando, come vedremo, un certo prezzo) solo passando al *campo dei numeri complessi*.

3.4 Alcune disuguaglianze

In Analisi si fa largo impiego di disuguaglianze; vogliamo raccoglierne in questo paragrafo alcune delle più importanti.

Dalla definizione di valore assoluto segue immediatamente che:

$$\forall a \geq 0\text{: } |x| \leq a \iff -a \leq x \leq a\ . \tag{3.6}$$

Dalla (3.6) segue l'importante *disuguaglianza triangolare*:

$$\forall x, y \in \mathbb{R}\text{: } |x+y| \leq |x| + |y|\ . \tag{3.7}$$

Infatti, basta scrivere le relazioni:

$$-|x| \leq x \leq |x| \quad \text{e} \quad -|y| \leq y \leq |y|$$

quindi sommare membro a membro, ottenendo

$$-(|x|+|y|) \leq x+y \leq |x|+|y|$$

da cui, per la (3.6), segue la (3.7).

La disuguaglianza triangolare è spesso usata nella forma seguente:

$$|a-b| \leq |a-c|+|b-c| \quad \forall a,b,c \in \mathbb{R} \ . \tag{3.8}$$

Basta porre nella (3.7) $x = a - c$ e $y = b - c$. Se invece, sempre nella (3.7), si pone $x = a - c$ e $y = b$ si ottiene

$$|a| \leq |a-b|+|b| \quad \text{cioè} \quad |a|-|b| \leq |a-b| \ .$$

Analogamente, scambiando a con b otteniamo, $|b| - |a| \leq |a - b|$ e dunque, ricordando la (3.6),

$$||a|-|b|| \leq |a-b| \ . \tag{3.9}$$

L'espressione $|a - b|$ rappresenta geometricamente la distanza (nel senso della geometria elementare) dei due punti a e b sulla retta.

La (3.7) può facilmente estendersi al caso di n addendi:

$$\left|\sum_{k=1}^{n} x_k\right| \leq \sum_{k=1}^{n} |x_k| \ . \tag{3.10}$$

La semplice dimostrazione si fa per induzione.

Ancora per induzione si prova la *disuguaglianza di Bernoulli*:

$$\forall x > -1 \ , \ \forall n \in \mathbb{N}\colon \ (1+x)^n \geq 1+nx \ . \tag{3.11}$$

Infatti la (3.11) è evidentemente vera per $n = 0$; se la riteniamo vera per un n generico, abbiamo

$$\begin{aligned}(1+x)^{n+1} &= (1+x)^n(1+x) \geq (1+nx)(1+x) = \\ &= 1+nx+x+nx^2 \geq 1+(n+1)x\end{aligned}$$

che prova la (3.11) per ogni n.

Un corollario della (3.11) è la seguente disuguaglianza,

$$\forall a > 1 \ , \ \forall n \in \mathbb{N}_+\colon \ 0 < \sqrt[n]{a}-1 \leq \frac{a-1}{n} \ . \tag{3.12}$$

Una forma equivalente di (3.12) è la seguente:

$$a \leq \left(1+\frac{a-1}{n}\right)^n \tag{3.12'}$$

già utilizzata nel par. 3.3. La (3.12') segue dalla (3.11) ponendo $x = (a-1)/n$.

La seguente disuguaglianza mostra che la *media geometrica* di due numeri (positivi) è minore della loro *media aritmetica*

$$\forall x, y \in \mathbb{R}_+ \ ; \ \sqrt{xy} \leq \frac{x+y}{2} \ . \tag{3.13}$$

La (3.13) si verifica immediatamente prendendo il quadrato di ambo i membri; si vede anche che il segno "=" vale se e solo se $x = y$.

La (3.13) si generalizza al caso di n numeri $x_1 \dots x_n$ reali positivi:

$$\sqrt[n]{x_1 x_2 \dots x_n} \leq \frac{(x_1 + x_2 + \cdots + x_n)}{n} \ . \tag{3.14}$$

La dimostrazione di (3.14) si può fare ancora utilizzando il principio di induzione.

Esercizi

1. Sia $x \in \mathbb{R}$. Quale relazione è vera?

a) $\sqrt{x^2} = x$

b) $\sqrt{x^2} = \pm x$

c) $\sqrt{x^2} = |x|$.

2. Se $x \in \mathbb{R}_+$, $k, m \in \mathbb{Z}$ e $n \in \mathbb{N}_+$, m, n primi tra loro, allora definiamo

$$x^{\frac{km}{kn}} := x^{\frac{m}{n}} \quad \text{ovvero} \quad \sqrt[kn]{x^{km}} := \sqrt[n]{x^m} \ . \tag{3.15}$$

Verificare che la (3.15) si estende a $x \in \mathbb{R}$ quando k ed n sono dispari.
Come si deve modificare la (3.15) per gli altri casi?

3. Commentare la validità dell'uguaglianza

$$\sqrt[n]{x^m} = (\sqrt[n]{x})^m \quad \text{ovvero} \quad (x^m)^{\frac{1}{n}} = (x^{\frac{1}{n}})^m \ .$$

4. Fareste dei calcoli per risolvere le seguenti disequazioni?

a) $\sqrt{x-2} < -1$

b) $\sqrt{-1-x^2} > 2x$.

5. Risolvere le disequazioni

a) $\sqrt{|x-1|} < 2 - x$

b) $\sqrt[3]{x^3+1} > x - 4$

c) $\dfrac{\sqrt{2x+1}-3}{x^2-4} \leq 0$.

6. Dimostrare le proprietà $E_0 \dots E_5$ degli esponenziali nel par. 3.2.

7. Dimostrare le proprietà $L_1, \dots, L_4$ dei logaritmi nel par. 3.3.

8. Quali relazioni sono vere? Sia $a > 0$, $a \neq 1$

a) $\forall x \in \mathbb{R}_+$, $x = a^{\log_a x}$,

b) $\forall x \in \mathbb{R}, x \neq 0, \log_a(x^2) = 2\log_a x$,

c) $\forall x \in \mathbb{R},\ x \neq 0,\ \log_a(x^2) = 2\log_a |x|$,

d) $\forall x, y \in \mathbb{R},\ xy > 0 \Longrightarrow \log_a(xy) = \log_a x + \log_a y$,

e) $\forall x, y \in \mathbb{R}, xy > 0 \Longrightarrow \log_a(xy) = \log_a |x| + \log_a |y|$.

9. Risolvere le seguenti disequazioni:

a) $\log_2(x-4) - \log_2(x-1) > 1$,

b) $(\log_2 x)^2 - 3\log_2 x - 5 \geq 0$.

10. Dimostrare che $\forall x, y \in \mathbb{R},\ 2xy \leq x^2 + y^2$. Quando vale l'uguaglianza?

11. Si dimostri che, $\forall x, y \in \mathbb{R},\ \forall \varepsilon > 0$

$$2xy \leq \varepsilon x^2 + \frac{1}{\varepsilon} y^2 .$$

4. I NUMERI COMPLESSI

Abbiamo già osservato che ci sono ragioni, di natura algebrica, che ci spingono ad ampliare ulteriormente il campo numerico: per esempio, l'osservazione, già altre volte fatta, che il polinomio di secondo grado $x^2 - 1$ ha in $\mathbb{R}$ precisamente due radici mentre l'analogo polinomio $x^2 + 1$ non ne ha alcuna (né in $\mathbb{R}$ né in alcun altro campo ordinato). D'altra parte sappiamo, per il teorema di isomorfismo (teorema 2.7), che nessun ampliamento sarà possibile se vorremo conservare sia la struttura algebrica che l'ordinamento. Sembra perciò naturale cercare una estensione di $\mathbb{R}$ ove sia possibile definire le solite operazioni algebriche e mantenerne le proprietà, sacrificando l'ordinamento.

4.1 Definizione di C e struttura di campo

Consideriamo $\mathbb{R}^2$, cioè l'insieme delle coppie ordinate (a, b) di numeri reali sulle quali definiamo direttamente le operazioni di somma e prodotto con le seguenti regole:

$$(a,b) + (c,d) := (a+c, b+d) \tag{4.1}$$

$$(a,b) \cdot (c,d) := (ac - bd, ad + bc) . \tag{4.2}$$

Osserviamo che, $\forall (a,b)$:

$$(a,b) + (0,0) = (0,0) + (a,b) = (a,b) ;$$

dunque la coppia $(0,0)$ è elemento neutro per la somma.

$$(a,b) \cdot (1,0) = (1,0) \cdot (a,b) = (a,b) ;$$

dunque la coppia $(1,0)$ è elemento neutro per il prodotto.

$$(a,b)+(-a,-b)=(0,0) ;$$

dunque $(-a,-b)$ è l'opposto di (a,b).
Se

$$(a,b)\neq(0,0) \ , \ (a,b)\cdot\left(\frac{a}{a^2+b^2},\frac{-b}{a^2+b^2}\right)=(1,0) ;$$

dunque la coppia $(a/(a^2+b^2),-b/(a^2+b^2))$ è l'inverso di (a,b).

Le proprietà usuali (commutativa, associativa, distributiva) sono verificate e perciò l'insieme $\mathbb{R}^2$ così strutturato è un campo, che chiamaremo *campo dei numeri complessi* e indicheremo con $\mathbb{C}$.

Osserviamo ora che $\mathbb{C}$ contiene il sottoinsieme $\mathbb{C}_0$ delle coppie del tipo $(a,0)$; esso è un *sottocampo* di $\mathbb{C}$, poiché somma e prodotto di coppie di questo tipo sono ancora coppie dello stesso tipo; infatti si ha:

$$(a,0)+(b,0)=(a+b,0)$$
$$(a,0)\cdot(b,0)=(ab,0) .$$

Inoltre $\mathbb{C}_0$ può essere ordinato ponendo $(a,0)<(b,0)$ se $a<b$.

Se allora mettiamo in corrispondenza biunivoca l'insieme dei numeri reali $\mathbb{R}$ con $\mathbb{C}_0$, ponendo

$$(a,0)\leftrightarrow a$$

si ottiene un *isomorfismo* tra *campi ordinati* e perciò possiamo *identificare* i numeri reali a con i numeri complessi del tipo $(a,0)$.

In questo senso il campo dei numeri complessi $\mathbb{C}$ è un ampliamento di quello dei numeri reali $\mathbb{R}$.

Consideriamo ora il numero $(0,1)$. Esso ha la singolare proprietà che:

$$(0,1)\cdot(0,1)=(-1,0)$$

cioè *il suo quadrato coincide col numero reale* -1*!* Per questa ragione la coppia $(0,1)$ merita di essere indicata con un simbolo speciale: la indicheremo con "i" e la chiameremo **unità immaginaria**.

A questo punto conviene semplificare la notazione. Osserviamo che, se scriviamo semplicemente a invece di $(a,0)$ etc., abbiamo

$$(a,b)=(a,0)+(0,1)\cdot(b,0)=a+ib .$$

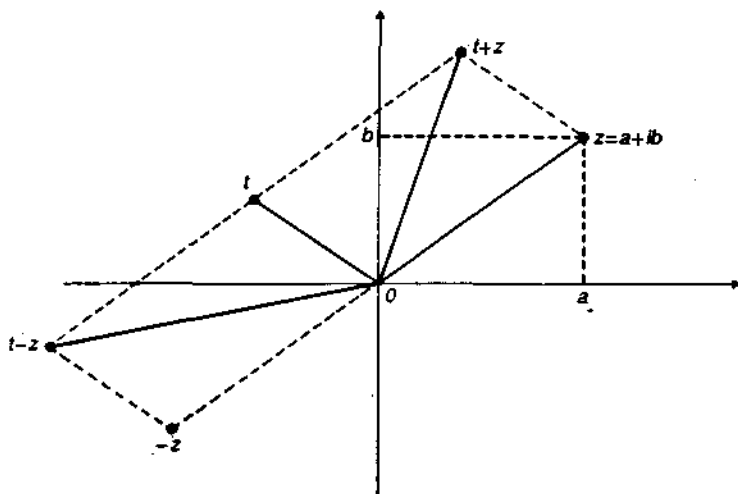



Fig. 2.2

Con questa notazione le regole (4.1) e (4.2) sono le ordinarie regole del calcolo letterale, ove si tenga conto che $i^2 = -1$:

$$(a + ib) + (c + id) = (a + c) + i(b + d)$$
$$(a + ib) \cdot (c + id) = (ac - bd) + i(ad + bc)$$

La scrittura

$$z = a + ib \tag{4.3}$$

è detta *forma cartesiana* o *algebrica* dei numeri complessi; a si chiama *parte reale* di z e si indica con $\mathrm{Re}(z)$, b si chiama *parte immaginaria* e si indica con $\mathrm{Im}(z)$.

I numeri complessi, essendo coppie di numeri reali, possono rappresentarsi come punti del piano (riferito ad assi cartesiani ortogonali): l'ascissa è la parte reale del numero, l'ordinata la parte immaginaria. In questo contesto il piano è detto *piano complesso* o *piano di Gauss*. In figura 2.2 sono illustrate geometricamente le operazioni di somma e differenza di due numeri complessi.

4.2 Coniugato, modulo e argomento

Il numero complesso $a - ib$ si dice il *complesso coniugato* di $z = a + ib$ e si indica con $\overline{z}$. Evidentemente si ha:

$$z + \overline{z} = 2\,\mathrm{Re}(z)$$
$$z - \overline{z} = 2i\,\mathrm{Im}(z)\ .$$

L'operazione di coniugio ha le seguenti elementari proprietà rispetto alla somma

e al prodotto:

$$\overline{(z_1 + z_2)} = \overline{z}_1 + \overline{z}_2$$
$$\overline{(z_1 z_2)} = \overline{z}_1 \cdot \overline{z}_2$$
$$\overline{\left(\frac{1}{z}\right)} = \frac{1}{\overline{z}} .$$

Inoltre questa operazione è *involutoria*, cioè: $\overline{\overline{z}} = z$. Pertanto $z\overline{z}$ è reale e precisamente, se $z = a + ib$, risulta

$$z\overline{z} = (a + ib)(a - ib) = a^2 + b^2 \geq 0 .$$

Si chiama *modulo* di $z = a + ib$ il numero reale non negativo $\sqrt{a^2 + b^2}$, che si indica con $|z|$. Se $z = a$ è reale, risulta $|z| = |a|$ (valore assoluto di a), in accordo con la precedente notazione.

Valgono le seguenti proprietà:

a) $|z| \geq 0$, $|z| = 0 \iff z = 0$

b) $|z| = |\overline{z}|$

c) $|\operatorname{Re}(z)| \leq |z|$, $|\operatorname{Im}(z)| \leq |z|$, $|z| \leq |\operatorname{Re}(z)| + |\operatorname{Im}(z)|$

d) $|z_1 z_2| = |z_1||z_2|$, $\left|\frac{1}{z}\right| = \frac{1}{|z|}$

e) $|z_1 + z_2| \leq |z_1| + |z_2|$ (disuguaglianza triangolare) (4.4)

f) $|z_1 + z_2| \geq ||z_1| - |z_2||$. (4.5)

Dimostriamo le (4.4) e (4.5).
Esse sono equivalenti alla seguente:

$$(|z_1| - |z_2|)^2 \leq |z_1 + z_2|^2 \leq (|z_1| + |z_2|)^2 .$$

Ponendo $z_1 = a + ib$, $z_2 = c + id$ otteniamo:

$$(\sqrt{a^2 + b^2} - \sqrt{c^2 + d^2})^2 \leq (a + c)^2 + (b + d)^2 \leq (\sqrt{a^2 + b^2} + \sqrt{c^2 + d^2})^2 .$$

Con calcoli elementari questa doppia disuguaglianza si riduce a

$$-\sqrt{a^2 + b^2} \cdot \sqrt{c^2 + d^2} \leq ac + bd \leq \sqrt{a^2 + b^2} \cdot \sqrt{c^2 + d^2} ,$$

che è equivalente alla seguente:

$$|ac + bd| \leq \sqrt{a^2 + b^2} \cdot \sqrt{c^2 + d^2} .$$

Elevando al quadrato entrambi i membri si arriva a:

$$(ac+bd)^2 \leq (a^2+b^2)(c^2+d^2)$$

ovvero a

$$0 \leq -2acbd + a^2d^2 + b^2c^2 = (ad-bc)^2$$

che è vera per ogni $a, b, c, d \in \mathbb{R}$. □

Geometricamente, $|z|$ rappresenta la *distanza* (nel senso usato in Geometria Euclidea) del punto (o numero complesso) z dall'origine; $|z_1 - z_2|$ rappresenta la distanza dei due punti z_1 e z_2; le disuguaglianze (4.4) e (4.5) traducono il noto teorema sulle lunghezze dei lati di un triangolo (v. fig. 2.3).

Dalla (4.4), ponendo $z_1 = \alpha - \gamma$ e $z_2 = \gamma - \beta$ otteniamo, $\forall \alpha, \beta, \gamma \in \mathbb{C}$

$$|\alpha - \beta| \leq |\alpha - \gamma| + |\gamma - \beta| \,. \tag{4.4'}$$

Come è noto dalla Geometria, i punti del piano possono essere individuati, oltre che dalle loro coordinate cartesiane, anche dalle coordinate polari: ρ (raggio polare, cioè distanza del punto dall'origine) e θ (angolo polare, cioè l'angolo che la retta congiungente il punto con l'origine forma con l'asse delle ascisse positive, contato in senso antiorario). È chiaro che una coppia ρ, θ, con $\rho > 0$, individua un ben determinato punto del piano; invece un punto del piano individua univocamente la coordinata ρ, ma l'angolo θ, misurato in radianti, è determinato solo a meno di multipli di 2π.

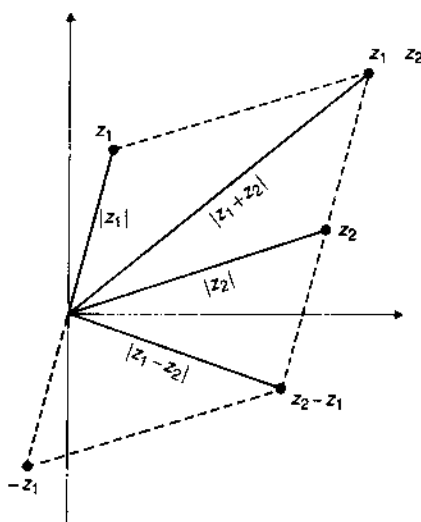



Fig. 2.3

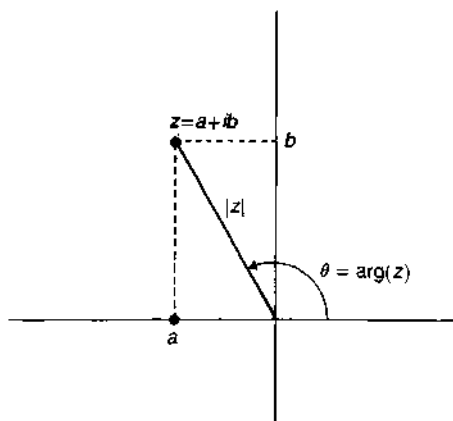



Fig. 2.4

Dato un numero complesso z, il suo modulo $|z|$ coincide col raggio polare del punto che ne è l'immagine sul piano complesso. Chiamiano *argomento* di z, e indicheremo con $\arg(z)$, uno qualsiasi degli angoli θ relativi al punto z. In questo modo l'argomento di z non è ben determinato. Spesso questa indeterminatezza non porta alcun inconveniente. Altre volte invece è preferibile assegnare un ben determinato argomento a un numero complesso. Ciò può ottenersi in infiniti modi, fissando un qualsiasi intervallo, di ampiezza 2π, entro il quale far variare l'angolo θ. Gli intervalli più comunemente usati a questo scopo sono $[0, 2\pi)$ e $(-\pi, \pi]$; allora l'argomento di z viene detto *argomento principale*.

***Esempio* 4.1** - Il numero $-i$ ha come argomento $-\pi/2$ oppure $3\pi/2$ oppure qualunque altro valore della forma $-\pi/2 + 2k\pi$ con $k \in \mathbb{Z}$. Il suo argomento principale sarà $3\pi/2$ se si adotta la convenzione che $\theta \in [0, 2\pi)$, $-\pi/2$ con la convenzione che $\theta \in (-\pi, \pi]$. I numeri reali positivi hanno argomento principale 0 e quelli negativi π con entrambe le convenzioni. Per il numero 0 l'argomento non viene definito.

Dato il numero $z = a + ib$, dalla trigonometria ricaviamo immediatamente le relazioni tra le coordinate cartesiane a, b e quelle polari ρ, θ:

$$a = \rho \cos\theta \ , \ b = \rho \sin\theta \tag{4.6}$$

cosicché il numero complesso z può anche scriversi nella forma

$$z = \rho(\cos\theta + i \sin\theta) \tag{4.7}$$

che è detta *forma trigonometrica* dei numeri complessi.

Le relazioni inverse di (4.6) sono:

$$\rho = \sqrt{a^2 + b^2} \ , \ \cos\theta = \frac{a}{\sqrt{a^2 + b^2}} \ , \ \sin\theta = \frac{b}{\sqrt{a^2 + b^2}} \ . \tag{4.8}$$

La forma trigonometrica è comoda per esprimere prodotti e quozienti di numeri complessi. Se abbiamo infatti

$$z_1 = \rho_1(\cos\theta_1 + i\sin\theta_1)\ ,\ \ z_2 = \rho_2(\cos\theta_2 + i\sin\theta_2)$$

otteniamo

$$\begin{aligned} z_1 z_2 &= \rho_1\rho_2\{\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2 + \\ &+ i(\sin\theta_1\cos\theta_2 + \cos\theta_1\sin\theta_2)\} = \\ &= \rho_1\rho_2\{\cos(\theta_1+\theta_2) + i\sin(\theta_1+\theta_2)\} \end{aligned} \tag{4.9}$$

e, se $z_2 \neq 0$,

$$\frac{z_1}{z_2} = \frac{\rho_1}{\rho_2}\cdot\frac{\cos\theta_1 + i\sin\theta_1}{\cos\theta_2 + i\sin\theta_2}\ ,$$

moltiplicando numeratore e denominatore per $\cos\theta_2 - i\sin\theta_2$ e tenendo conto che $(\cos\theta_2)^2 + (\sin\theta_2)^2 = 1$ abbiamo:

$$\begin{aligned} \frac{z_1}{z_2} &= \frac{\rho_1}{\rho_2}(\cos\theta_1 + i\sin\theta_1)(\cos\theta_2 - i\sin\theta_2) = \\ &= \frac{\rho_1}{\rho_2}\{\cos(\theta_1-\theta_2) + i\sin(\theta_1-\theta_2)\}\ . \end{aligned} \tag{4.10}$$

Pertanto il modulo del prodotto e del quoziente di due numeri complessi è, rispettivamente, il prodotto e il quoziente dei moduli; l'argomento è la somma e, rispettivamente, la differenza degli argomenti.

4.3 Potenze e radici

La (4.9) si generalizza (per induzione) al caso di un numero qualsiasi di fattori $z_1, z_2, \dots, z_n$.

$$\begin{aligned} z_1 z_2 \dots z_n = \rho_1\rho_2\dots\rho_n\{&\cos(\theta_1+\theta_2+\dots+\theta_n) + \\ &+ i\sin(\theta_1+\theta_2+\dots+\theta_n)\}\ . \end{aligned} \tag{4.11}$$

Se poi i fattori sono tutti uguali, otteniamo la formula (detta di De Moivre):

$$z^n = \rho^n(\cos(n\theta) + i\sin(n\theta))\ . \tag{4.12}$$

Dato un numero complesso w, diremo che z è una *radice n-esima (complessa)* di w se risulta $z^n = w$.

▪ **Teorema 4.1** - *Sia $w \in \mathbb{C}$ e n intero ≥ 1. Esistono precisamente n radici n-esime complesse $z_0, z_1, \dots, z_{n-1}$ di w; posto $w = r(\cos\varphi + i\sin\varphi)$ e $z_k = \rho_k(\cos\theta_k + i\sin\theta_k)$ abbiamo*

$$\rho_k = r^{\frac{1}{n}}\ ,\ \ \theta_k = \frac{\varphi + 2k\pi}{n} \quad k = 0, 1, \dots, n-1\ . \tag{4.13}$$

Dimostrazione - Se $w = 0$, l'equazione $z^n = 0$ ha n radici coincidenti in 0 e il teorema è verificato. Sia $w \neq 0$; i numeri z_k sono evidentemente radici di w, come risulta applicando la formula di De Moivre. Mostriamo che non ve ne sono altre. Infatti, perché un numero $R(\cos\psi + i\sin\psi)$ sia radice n-esima di w, dovrebbe risultare

$$R^n = r \quad \text{e} \quad n\psi = \varphi + 2h\pi \quad \text{con } h \in \mathbb{Z}$$

e cioè

$$R = r^{1/n} \quad \text{e} \quad \psi = \varphi/n + 2h\pi/n\ .$$

Dando ad h i valori $0, 1, \ldots, n-1$ troviamo appunto i numeri z_k. Dando ad h un qualsiasi altro valore $\overline{h}$ diverso dai precedenti, questo può scriversi nella forma $\overline{h} = k + mn$ ($m \in \mathbb{Z}$ è il quoziente e k è il resto della divisione di $\overline{h}$ per n) per cui sarebbe

$$\psi = \frac{\varphi}{n} + \frac{2k\pi}{n} + 2m\pi = \theta_k + 2m\pi$$

e ritroveremmo ancora gli stessi z_k precedenti. □

Osservazione 4.1 - Per le radici complesse si usa purtroppo una notazione un po' ambigua, la stessa in uso per indicare la radice aritmetica; si indica cioè con $\sqrt[n]{z}$ o $z^{1/n}$ l'insieme delle n radici complesse di z (oppure una in particolare di esse). Ciò può creare confusione quando z è reale. Infatti il simbolo $\sqrt{4}$, inteso come radice aritmetica di 4, è 2; inteso come radice complessa di 4 è l'insieme dei due numeri, $+2$ e -2.

***Esempio* 4.2** - Le 3 radici cubiche di -1 sono i numeri z_k della forma: $\cos\theta_k + i\sin\theta_k$ con $\theta_k = \pi/3 + 2k\pi/3$, $k = 0, 1, 2$.

Esplicitamente abbiamo:

$$z_0 = \frac{1}{2}(1 + \sqrt{3}i)$$

$$z_1 = -1$$

$$z_2 = \frac{1}{2}(1 - \sqrt{3}i)$$

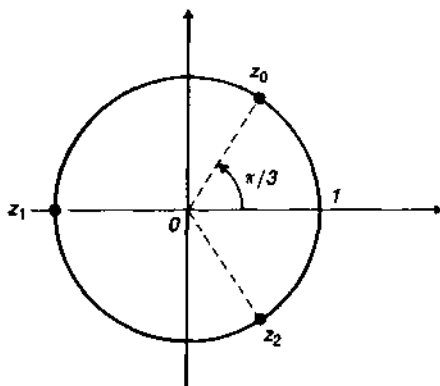



Fig. 2.5

Osservazione 4.2 - La disposizione delle radici cubiche dell'esempio precedente nel piano di Gauss non è casuale.

Infatti se $w = r(\cos\varphi + i\sin\varphi)$ le radici n-esime $z_0, z_1, \ldots, z_{n-1}$ di w si trovano ai vertici del poligono regolare di n-lati inscritto nella circonferenza di centro 0 e raggio $r^{1/n}$, con il vertice z_0 posto nel punto di argomento $\theta = \varphi/n$.

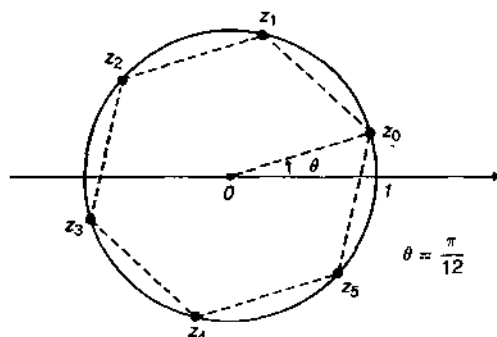



Fig. 2.6

Nella figura 2.6 sono rappresentate le radici seste di i: $z_0, z_1, \ldots, z_5$.

Il teorema 4.1 ci dice non solo che il polinomio $x^2 + 1$ ha in $\mathbb{C}$ esattamente due radici, ma, più in generale, che un polinomio del tipo $x^n + a$ (con a complesso) ha in $\mathbb{C}$ esattamente n radici. L'anomalia incontrata nel campo reale circa il numero delle radici dell'equazione $x^n = y$ è ora chiarita: tale equazione ha sempre n radici in $\mathbb{C}$, ma solo occasionalmente una o due di queste stanno in $\mathbb{R}$. Il risultato è di portata ben più generale, come afferma il seguente teorema, di cui non riportiamo la dimostrazione.

▪ **Teorema 4.2** - (teorema fondamentale dell'algebra). *Una equazione polinomiale della forma*

$$a_0 + a_1 x + \cdots + a_n x^n = 0 \qquad (a_n \neq 0)$$

con coefficienti complessi qualsiansi ha precisamente n radici in $\mathbb{C}$, se ognuna di esse viene contata con la sua molteplicità (*).

Questo teorema mostra chiaramente come il campo dei numeri complessi sia quello più idoneo per trattare le equazioni algebriche. Più avanti definiremo sui numeri complessi anche l'operazione di elevamento a potenza con esponente complesso e il logaritmo, ottenendo una soddisfacente sistemazione per l'insieme di tutte queste operazioni.

Col campo complesso abbiamo concluso la descrizione dei campi numerici. Potrebbe nascere la curiosità di sapere quali ulteriori ampliamenti siano possibili;

(*) Se $P(z)$ è un polinomio in z di grado n e z_0 una sua radice, si dice che z_0 è di molteplicità k (k intero, $k \geq 1$) se vale la formula $P(z) = (z - z_0)^k Q(z)$ dove Q è un polinomio tale che $Q(z_0) \neq 0$. Per maggiori notizie sui polinomi, v. Cap. 5, sez. 3.

ma tali generalizzazioni non sono di interesse per l'Analisi, poiché, per procedere oltre, bisognerebbe rinunciare a qualcuna di quelle proprietà fondamentali che caratterizzano la struttura algebrica dei corpi.

Esercizi

1. Disegnare nel piano complesso il luogo dei punti $z = (\rho, \theta)$ tali che:

a) $\rho = 3$

b) $\theta = \dfrac{\pi}{3}$

c) $\mathrm{Re}(z) \geq -1$ e $\mathrm{Im}(z) < 3$.

2. Disegnare nel piano complesso il luogo dei punti z tali che:

a) $|z| = |z + i|$,

b) $\mathrm{Re}(z^2) > k$ $(k \in \mathbb{R})$,

c) $\mathrm{Im}\,\dfrac{1}{z} = k$ $(k \in \mathbb{R})$.

3. Verificare che, per ogni $z \in \mathbb{C}$:

a) $|z| \geq \mathrm{Re}(z)$,

b) $|z| \geq \mathrm{Im}(z)$,

c) $|z| \leq |\mathrm{Re}(z)| + |\mathrm{Im}(z)|$.

4. Dimostrare che l'area del triangolo di vertici $0, z, w$ è $\dfrac{1}{2}\,|\mathrm{Im}(\overline{z}w)|$.

5. Scrivere in forma cartesiana $a + ib$ i numeri:

$$(1+i)^{20} \quad , \quad (1-i)^{11} \, .$$

6. Calcolare le radici seste di -1 e rappresentarle sul piano di Gauss.

7. Risolvere l'equazione $(z-2)^3 = -i$.

8. Dimostrare che l'equazione di una retta nel piano complesso si può scrivere nella forma:

$$az + \overline{a}\,\overline{z} + c = 0 \quad \text{con} \quad a \in \mathbb{C} \text{ e } c \in \mathbb{R} \, .$$

9. Dimostrare che l'equazione di una circonferenza nel piano complesso ha la forma:

$$az\overline{z} + bz + \overline{b}\overline{z} + c = 0 \quad \text{con} \quad a, b \in \mathbb{C} \text{ e } c \in \mathbb{R} \, .$$

A quale condizione devono soddisfare a, b, c affinché la circonferenza sia reale?

10. Studiare l'effetto della trasformazione del piano complesso in sé data da $z \to 1/z$. Dimostrare che trasforma rette o cerchi in rette o cerchi (non rispettivamente).

11. Dimostrare che l'effetto della moltiplicazione per un numero complesso z di modulo unitario ed argomento θ è una rotazione di centro origine ed angolo θ.

12.* Nell'insieme dei numeri complessi introduciamo l'ordinamento lessicografico seguente.

Dati $z = a + ib$ e $w = c + id$, diciamo che $z < w$ se $a < c$ oppure se $a = c$ e $b < d$.

Verificare che l'ordinamento così introdotto è *incompatibile* con la struttura algebrica di $\mathbb{C}$.

13.* Dimostrare che *ogni* possibile ordinamento totale introdotto nell'insieme dei numeri complessi è incompatibile con la struttura algebrica di $\mathbb{C}$.

Risoluzione delle equazioni di 3° grado (Scipione dal Ferro (1515), Girolamo Cardano (1545)).

Ogni equazione di 3° grado: $y^3 + by^2 + cy + d = 0$ può scriversi nella forma

$$x^3 + px + q = 0 \tag{4.14}$$

(basta porre $y = x - b/3$). Cerchiamo ora u, v tali che una soluzione della (4.14) abbia la forma $u + v$; si trova subito che deve essere

$$uv = -\frac{p}{3} \qquad u^3 + v^3 = -q$$

e perciò, elevando al cubo la prima equazione, si trova che u^3 e v^3 sono le radici dell'equazione di 2° grado: $z^2 + qz - p^3/27 = 0$ cioè

$$u^3 = -\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}} \qquad v^3 = -\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}$$

e infine

$$x = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}\,. \tag{4.15}$$

I numeri rappresentati dalla formula (4.15) sono 9, poiché ogni radice cubica ha tre valori; ma quelli che risolvono la (4.14) sono soltanto tre, perché deve essere $uv = -p/3$; perciò, se $u_k (k = 1, 2, 3)$ sono i tre valori del primo radicale, le tre soluzioni di (4.14) saranno:

$$x_k = u_k - \frac{p}{3u_k} \quad (k = 1, 2, 3)\,.$$

14. Mostrare che, se p e $q \in \mathbb{R}$, l'equazione (4.14) ha sempre almeno una radice reale; quando tutte e tre le radici sono reali?

15. Risolvere le equazioni:

$$x^3 - 3x - 4 = 0 \quad , \quad x^3 - 9x + 10 = 0\,.$$

Spazi euclidei 3

In questo capitolo presentiamo le principali strutture dell'insieme
$\underbrace{\mathbb{R} \times \mathbb{R} \times \ldots \times \mathbb{R}}_{n \text{ volte}} = \mathbb{R}^n$ (e analogamente di $\mathbb{C}^n = \mathbb{C} \times \mathbb{C} \times \ldots \times \mathbb{C}$).

La prima è di tipo algebrico: quella di *spazio lineare* o *vettoriale* che si rivelerà fondamentale nel seguito e non solo per i corsi di Analisi.

Nell'ambito di questa struttura si precisa il concetto di *dimensione*; si introducono poi i concetti di *prodotto scalare*, di *norma*, di *distanza* e di *topologia*.

La topologia di $\mathbb{R}^n$ e dei suoi possibili ampliamenti ottenuti aggiungendo "i punti all'infinito" è trattata nei suoi aspetti principali nella sezione 2. Come vedremo nel capitolo 4 la nozione di limite è un *concetto topologico*: su tale nozione si appoggerà poi il concetto di derivata e tutto il susseguente sviluppo del calcolo differenziale.

1. GLI SPAZI EUCLIDEI: $\mathbb{R}^n$ E $\mathbb{C}^n$

1.1 Spazi vettoriali lineari

Nel piano (nello spazio fisico) riferito a un sistema di assi cartesiani ortogonali, i punti sono rappresentati da coppie (x_1, x_2) (terne (x_1, x_2, x_3)) di numeri reali, che sono le loro coordinate; si possono anche pensare come il secondo estremo P di segmenti orientati aventi il primo estremo nell'origine O degli assi coordinati.

Sull'insieme di questi punti sappiamo fare due operazioni fondamentali: la somma (secondo la usuale regola del parallelogramma) e il prodotto per un numero, illustrate in figura 3.1.

Se P e Q (più precisamente: i segmenti OP e OQ) rappresentano rispettivamente due forze applicate in O, allora $P + Q$ rappresenta la loro risultante, $2P$ rappresenta una forza, pure applicata in O, con la stessa direzione e verso di P ma con intensità doppia, $-\frac{1}{2}P$ rappresenta una forza con la stessa direzione di P, verso contrario e intensità ridotta di un mezzo.

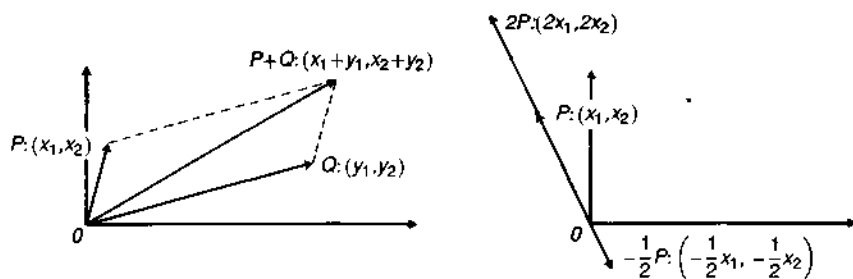



Fig. 3.1

La possibilità di poter eseguire sui punti queste operazioni conferisce al loro insieme (indipendentemente dal particolare riferimento adottato) una speciale struttura, che ora illustreremo, così come la possibilità di eseguire sui numeri (razionali, reali, complessi) le operazioni di somma e prodotto (con le note proprietà) conferiva loro la struttura algebrica di campo.

La struttura che ora vogliamo esaminare è quella di *spazio lineare*. Per costruire uno spazio lineare servono: un insieme X e un campo Λ; quest'ultimo sarà per noi uno dei campi numerici precedentemente introdotti: o $\mathbb{R}$ o $\mathbb{C}$ (più raramente, $\mathbb{Q}$).

Definizione 1.1 - *Un insieme X si dice* **spazio vettoriale lineare** *sul campo Λ se sui suoi elementi sono definite due operazioni con le proprietà sottoelencate:*

$\mathcal{V}_1$. *È definita un'applicazione: $X \times X \to X$, che si indica col segno "$+$" e si chiama somma, in modo che:*

1. $\forall \mathbf{x}, \mathbf{y} \in X : \mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$
2. $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in X : (\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$
3. *esiste in X un elemento (elemento neutro, indicato con* $\mathbf{0}$*) tale che:* $\forall \mathbf{x} \in X : \mathbf{x} + \mathbf{0} = \mathbf{x}$
4. $\forall \mathbf{x} \in X$, *esiste un elemento (l'opposto di* $\mathbf{x}$, *indicato con* $-\mathbf{x}$*) tale che:* $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$

$\mathcal{V}_2$. *È definita un'applicazione: $\Lambda \times X \to X$ (moltiplicazione per gli elementi di Λ), indicata con $\lambda\mathbf{x}$, se $\lambda \in \Lambda$ e $\mathbf{x} \in X$, tale che,* $\forall \lambda, \mu \in \Lambda$, $\forall \mathbf{x}, \mathbf{y} \in X$:

1. $\lambda(\mu\mathbf{x}) = (\lambda\mu)\mathbf{x}$
2. $1\mathbf{x} = \mathbf{x}$
3. $(\lambda + \mu)\mathbf{x} = \lambda\mathbf{x} + \mu\mathbf{x}$
4. $\lambda(\mathbf{x} + \mathbf{y}) = \lambda\mathbf{x} + \lambda\mathbf{y}$.

Gli elementi di X si chiamano **vettori** o **punti** e si indicheranno (generalmente) con lettere latine minuscole in grassetto: $\mathbf{x}$, $\mathbf{y}$, $\mathbf{a}, \ldots$; gli elementi del campo si chiamano **scalari** e si indicheranno con lettere (greche o latine) minuscole in corsivo: $\alpha, \lambda, \mu, \ldots, x_1, x_2, \ldots$. Non si confonda l'elemento neutro di X (lo zero o origine dello spazio), indicato con $\mathbf{0}$ con lo zero del campo, indicato con 0.

Dalle proprietà $\mathcal{V}_1$ e $\mathcal{V}_2$ seguono poi facilmente le seguenti (che ogni studente potrà provare da sé)

4. $0 \cdot \mathbf{x} = \mathbf{0}$
5. $(-1)\mathbf{x} = -\mathbf{x}$.

cioè, moltiplicando per il numero 0 un qualsiasi vettore **x** si ottiene il vettore nullo, e l'opposto del vettore **x** si ottiene moltiplicando il vettore stesso per lo scalare -1.

Si dimostra anche che lo zero dello spazio e l'opposto di ogni elemento sono unici.

Ancora qualche definizione di carattere generale.

Definizione 1.2 - *Si dice* **combinazione lineare** *dei vettori* $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^k$ *un vettore* **x** *della forma*

$$\mathbf{x} = \lambda_1 \mathbf{x}^1 + \lambda_2 \mathbf{x}^2 + \cdots + \lambda_k \mathbf{x}^k$$

con $\lambda_1, \lambda_2, \ldots, \lambda_k$ scalari, detti *coefficienti* della combinazione.

Definizione 1.3 - *Si dice che k vettori* $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^k$ *sono* **linearmente dipendenti** *se esistono k scalari non tutti nulli,* $\alpha_1, \alpha_2, \ldots, \alpha_k$ *tali che*

$$\alpha_1 \mathbf{x}^1 + \alpha_2 \mathbf{x}^2 + \cdots + \alpha_k \mathbf{x}^k = \mathbf{0}\ . \tag{1.1}$$

In caso contrario si dicono **linearmente indipendenti.**

Notiamo che se i k vettori $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^k$ sono linearmente *dipendenti* allora *almeno uno* di essi è combinazione lineare degli altri.

Infatti, se vale la (1.1) con $\alpha_1 \neq 0$, per esempio, allora si deduce che

$$\mathbf{x}^1 = -\frac{\alpha_2}{\alpha_1}\mathbf{x}^2 - \frac{\alpha_3}{\alpha_1}\mathbf{x}^3 - \cdots - \frac{\alpha_k}{\alpha_1}\mathbf{x}^k$$

e cioè che $\mathbf{x}^1$ è combinazione lineare degli altri.

Viceversa se i k vettori sono *indipendenti* allora è chiaro che $\lambda_1 \mathbf{x}^1 + \lambda_2 \mathbf{x}^2 + \cdots + \lambda_k \mathbf{x}^k = \mathbf{0}$ *se* e *solo se* $\lambda_1 = \lambda_2 = \ldots = \lambda_k = 0$.

Definizione 1.4 - *Lo spazio X si dice* **finito-dimensionale** *se esiste un sottoinsieme finito* $B \subseteq X$ *con le seguenti proprietà:*

i) gli elementi di B sono linearmente indipendenti

ii) ogni vettore di X si può esprimere come combinazione lineare degli elementi di B.

Si dice anche che B è una **base** per X. Si dimostra (in Algebra) che tutte le basi hanno lo stesso numero cardinale: questo numero si chiama **dimensione** dello spazio.

Definizione 1.5 - *In uno spazio di dimensione n, si prendano k vettori; si chiama* **sottospazio** *generato dai k vettori l'insieme di tutte le loro combinazioni lineari.*
Se i k vettori sono indipendenti (e allora deve essere $k \leq n$) il sottospazio ha dimensione k.

Il presente capitolo è dedicato allo studio degli spazi lineari a dimensione finita. I concetti di base e di dimensione si estendono anche al caso in cui i vettori (componenti la base) sono infiniti. Lo studente incontrerà più avanti, nel corso dei suoi studi, esempi di tali spazi.

1.2 Gli spazi $\mathbb{R}^n$ e $\mathbb{C}^n$

L'insieme $\mathbb{R}^n$ (cioè il prodotto cartesiano $\mathbb{R} \times \mathbb{R} \times \ldots \times \mathbb{R}$ n volte, i cui elementi sono le n-uple ordinate di numeri reali) possiede in maniera naturale la struttura di spazio vettoriale.
Sia $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$. Il numero reale x_i $(1 \leq i \leq n)$ viene detto *componente* i-esima del vettore $\mathbf{x}$ o anche *coordinata* i-esima del punto $\mathbf{x}$.
Dato un numero reale λ si definisce il prodotto $\lambda\mathbf{x}$ ponendo

$$\lambda\mathbf{x} := (\lambda x_1, \lambda x_2, \ldots, \lambda x_n) \tag{1.2}$$

e, dato un altro vettore $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, si definisce la somma $\mathbf{x} + \mathbf{y}$ ponendo

$$\mathbf{x} + \mathbf{y} := (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n) \ . \tag{1.3}$$

Le proprietà $\mathcal{V}_1$ e $\mathcal{V}_2$ sono di immediata verifica; in particolare sarà

$$\mathbf{0} = (0, 0, \ldots, 0)$$
$$-\mathbf{x} = (-x_1, -x_2, \ldots, -x_n) \ .$$

L'insieme $\mathbb{R}^n$ così strutturato diventa quindi uno spazio (vettoriale) lineare sul campo $\mathbb{R}$.

***Esempio* 1.1** - Il vettore $\mathbf{x} = (7, 3, 6) \in \mathbb{R}^3$ è combinazione lineare dei vettori $\mathbf{x}^1 = (2, 0, 1)$ e $\mathbf{x}^2 = (-1, 3, 2)$ con coefficienti $\lambda_1 = 4$ e $\lambda_2 = 1$; infatti risulta

$$4(2, 0, 1) + (-1, 3, 2) = (7, 3, 6) \ .$$

***Esempio* 1.2** - Nello spazio fisico $\mathbb{R}^3$ consideriamo un sistema di k masse puntiformi $m_1, m_2, \ldots, m_k$ collocate nei punti $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^k$ rispettivamente. Il baricentro $\mathbf{b}$ del sistema è dato da:

$$\mathbf{b} = \frac{m_1\mathbf{x}^1 + m_2\mathbf{x}^2 + \cdots + m_k\mathbf{x}^k}{M}$$

dove si è posto $M = m_1 + m_2 + \cdots + m_k$ (massa totale del sistema). Perciò $\mathbf{b}$ è combinazione lineare dei vettori $\mathbf{x}^i$ con coefficienti $\lambda_i = m_i/M$ $(i = 1, 2, \ldots, k)$.

Si osservi che i coefficienti λ_i della combinazione sono tutti *non negativi* e tali che: $\lambda_1 + \lambda_2 + \cdots + \lambda_n = 1$; quando questo succede, la combinazione lineare si dice *convessa*.

***Esempio* 1.3** - I vettori di $\mathbb{R}^4$: $\mathbf{x}^1 = (0, 1, 0, 1)$, $\mathbf{x}^2 = (1, 0, 1, 0)$, $\mathbf{x}^3 = (1, 2, 2, 1)$ sono linearmente indipendenti; infatti l'equazione

$$\lambda_1\mathbf{x}^1 + \lambda_2\mathbf{x}^2 + \lambda_3\mathbf{x}^3 = \mathbf{0}$$

significa, scrivendo componente per componente, che $\lambda_1, \lambda_2, \lambda_3$, sono soluzioni del sistema lineare:

$$\begin{aligned} \lambda_2 + \lambda_3 &= 0 \\ \lambda_1 + 2\lambda_3 &= 0 \\ \lambda_2 + 2\lambda_3 &= 0 \\ \lambda_1 + \lambda_3 &= 0 \end{aligned}$$

che ha, come unica soluzione, $\lambda_1 = \lambda_2 = \lambda_3 = 0$.

Consideriamo ora i seguenti vettori di $\mathbb{R}^n$:

$$\begin{aligned} \mathbf{e}^1 &:= (1,0,0,\ldots,0) \\ \mathbf{e}^2 &:= (0,1,0,\ldots,0) \\ &\vdots \\ \mathbf{e}^n &:= (0,0,0,\ldots,1)\ . \end{aligned} \tag{1.4}$$

È chiaro che essi sono linearmente indipendenti; inoltre, dato un qualsiasi altro vettore $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ si può scrivere

$$\mathbf{x} = x_1\mathbf{e}^1 + x_2\mathbf{e}^2 + \cdots + x_n\mathbf{e}^n$$

e quindi questi vettori costituiscono una base per $\mathbb{R}^n$ detta *base canonica* (o sistema di assi cartesiani ortogonali). $\mathbb{R}^n$ è perciò uno spazio n-dimensionale. Per esempio, in $\mathbb{R}^3$ la base canonica è costituita dai 3 vettori che in Meccanica e in Fisica vengono di solito indicati con le lettere $\mathbf{i} := (1,0,0)$, $\mathbf{j} := (0,1,0)$ e $\mathbf{k} := (0,0,1)$. Un'altra base è costituita, ad esempio, dai 3 vettori $\mathbf{a}_1 = (0,1,1)$, $\mathbf{a}_2 = (1,0,1)$, $\mathbf{a}_3 = (1,1,0)$.

Osservazione 1.1 - Quando si scrive "il vettore $\mathbf{x} = (x_1, \ldots, x_n)$" si sottointende che, come *base* in $\mathbb{R}^n$ si è scelta quella canonica; dal punto di vista geometrico ciò significa introdurre, ad esempio in $\mathbb{R}^3$, una terna di assi cartesiani ortogonali come un sistema di riferimento nello spazio.

Non sempre la base canonica è la più appropriata ovvero non sempre il riferimento cartesiano ortogonale è il più conveniente; si può allora *cambiare base*; ad esempio in $\mathbb{R}^2$ nessuno vieta di scegliere come base, anziché $\mathbf{e}^1, \mathbf{e}^2$ il sistema di vettori

$$\mathbf{a}^1 = \left(1, \frac{1}{2}\right)\ ,\ \mathbf{a}^2 = (-1, 2)\ .$$

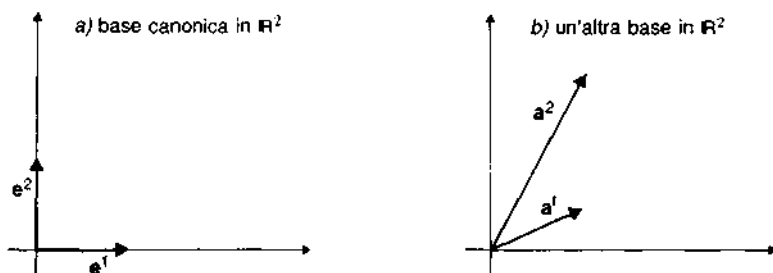



Fig. 3.2

In $\mathbb{R}^n$ k vettori linearmente indipendenti, $k < n$, generano sottospazi di dimensione k.

Esempi

1.4. In $\mathbb{R}^3$ sia $\mathbf{x} = (1, 1, 1)$. Il sottospazio (unidimensionale) V generato da $\mathbf{x}$ è l'insieme dei punti del tipo $\lambda\mathbf{x}$, $\lambda \in \mathbb{R}$; esso è la retta indicata in figura 3.3.

1.5. Sempre in $\mathbb{R}^3$ siano $\mathbf{x} = (1, 1, 1)$ e $\mathbf{y} = (1, 1, 0)$. Il sottospazio V generato da $\mathbf{x}$ e $\mathbf{y}$ è l'insieme dei punti del tipo

$$\lambda\mathbf{x} + \mu\mathbf{y} .$$

Tali punti descrivono il piano indicato in figura 3.4 passante per l'origine e contenente i vettori $\mathbf{x}$ e $\mathbf{y}$.

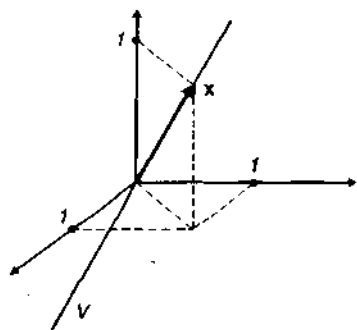



Fig. 3.3

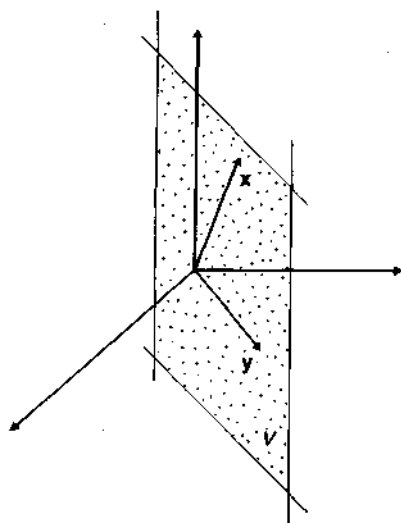



Fig. 3.4

Osservazione 1.2 - Si noti che, nella definizione 1.1 di spazio lineare, non si esclude il caso che X coincida con Λ (adottando come somma di vettori e prodotto di vettore per uno scalare le stesse espressioni già definite su Λ); cosicché gli elementi di Λ hanno il duplice aspetto di vettori e di scalari. È ciò che accade per l'insieme $\mathbb{R}^1 = \mathbb{R}$ che ha la struttura di spazio vettoriale (ad una dimensione) sul campo $\mathbb{R}$.

Ciò che è stato fatto per $\mathbb{R}^n$ immediatamente si estende all'insieme $\mathbb{C}^n$ (n-uple ordinate di numeri complessi).

Ora il campo Λ è quello dei complessi, la definizione di somma di vettori e prodotto di vettore per scalare è la stessa precedentemente introdotta. I vettori (1.4) sono ancora una base per $\mathbb{C}^n$ (poiché il vettore $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ si può sempre scrivere come $x_1\mathbf{e}^1 + x_2\mathbf{e}^2 + \cdots + x_n\mathbf{e}^n$) che risulta pertanto anch'esso uno spazio n-dimensionale. $\mathbb{C}$ stesso è uno spazio unidimensionale su $\mathbb{C}$.

Osservazione 1.3 - Dal punto di vista insiemistico $\mathbb{C}$ e $\mathbb{R}^2$ coincidono (cioè hanno gli stessi elementi, che sono le coppie ordinate di numeri reali); non così dal punto di vista della struttura algebrica, poiché $\mathbb{C}$ ha la struttura di campo e, se vogliamo, quella di spazio lineare unidimensionale su $\mathbb{C}$, mentre $\mathbb{R}^2$ ha la struttura di spazio lineare bidimensionale sul campo $\mathbb{R}$.

Osservazione 1.4 - I due esempi sopra illustrati di spazi n-dimensionali lineari sul campo $\mathbb{R}$ e, rispettivamente, $\mathbb{C}$, sono, sostanzialmente, gli *unici* possibili. Infatti, sia X uno spazio lineare n-dimensionale su $\mathbb{R}$ e sia $\{\mathbf{a}^1, \mathbf{a}^2, \ldots, \mathbf{a}^n\}$ una base per X. Ogni vettore $\mathbf{x} \in X$ si può allora rappresentare (in un unico modo!) nella forma: $\lambda_1\mathbf{a}^1 + \lambda_2\mathbf{a}^2 + \cdots + \lambda_n\mathbf{a}^n$; cioè ad $\mathbf{x}$ viene univocamente associato l'elemento $(\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n$. Viceversa, ad ogni elemento di $\mathbb{R}^n$ resta associato un ben determinato $\mathbf{x} \in X: (\lambda_1, \lambda_2, \ldots, \lambda_n) \to \lambda_1\mathbf{a}^1 + \lambda_2\mathbf{a}^2 + \cdots + \lambda_n\mathbf{a}^n$. Perciò la corrispondenza tra X e $\mathbb{R}^n$ è biunivoca e, come subito si accerta, conserva le operazioni algebriche; in altre parole X e $\mathbb{R}^n$ sono *isomorfi*.

Le stesse cose diconsi per $\mathbb{C}^n$ ed uno spazio n-dimensionale su $\mathbb{C}$.

Gli spazi $\mathbb{R}^n$ e $\mathbb{C}^n$ ora introdotti, così come gli spazi ad essi isomorfi, diconsi euclidei (*). In questa accezione, il termine "euclideo" è sinonimo di "finito dimensionale".

*1.3 Gruppi

Le due strutture algebriche che lo studente ha finora incontrato, quella di campo (o corpo commutativo) e quella di spazio lineare sono basate entrambe su *due* operazioni: tutt'e due di composizione interna (somma e prodotto di scalari) per gli elementi del campo, una interna e una esterna (somma di vettori e prodotto

(*) Talvolta il termine "euclideo" è riservato allo spazio $\mathbb{R}^n$ e suoi isomorfi, mentre per $\mathbb{C}^n$ si usa il termine "hermitiano".

di un vettore per uno scalare) per gli elementi dello spazio lineare. Alla base di entrambe queste strutture ce n'è una più semplice, basata su *una sola* operazione: la struttura di **gruppo**, che ora definiamo.

Sia X un insieme e sia definita un'applicazione $X \times X \to X$ che associa, ad ogni coppia $(x, y) \in X \times X$, un elemento di X che possiamo provvisoriamente indicare con $x * y$.

Definizione 1.6 - *L'insieme X si dice* **gruppo** *rispetto all'operazione definita se valgono le seguenti proprietà:*

$\mathcal{G}$ 1. $\forall a, b, c \in X: (a * b) * c = a * (b * c)$ (associativa)

2. *esiste un elemento (elemento neutro, indicato con e) tale che:*

$\forall a \in X: a * e = e * a = a$

3. $\forall a \in X$ *esiste un elemento (l'inverso di a, indicato con a^{-1}) tale che:*

$a * a^{-1} = a^{-1} * a = e.$

Dalle 1., 2., 3. discende poi che in X esiste un solo elemento neutro e che l'inverso di un dato elemento è unico. Si noti che la proprietà commutativa non è richiesta; se essa vale, cioè se

4. $\forall a, b \in X: a * b = b * a$

il gruppo si dice *commutativo* (o *abeliano*).

L'operazione definita in X potrà, a seconda dei casi, chiamarsi *somma* (e allora il gruppo si dirà additivo e si useranno le notazioni consuete: $a + b$ in luogo di $a * b$, 0 per l'elemento neutro, $-a$ per l'elemento inverso di a, etc.) oppure *prodotto* (e scriveremo ab in luogo di $a * b$ etc.) oppure, per esempio, prodotto di *composizione* (e potremo scrivere $a \circ b$ in luogo di $a * b$ ...) e così via.

Esempi di gruppi:

1.6. $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{C}$ sono gruppi commutativi rispetto all'addizione in essi definita.

1.7. $\mathbb{Q} \setminus \{0\}$, $\mathbb{R} \setminus \{0\}$, $\mathbb{C} \setminus \{0\}$ sono gruppi commutativi rispetto alla moltiplicazione in essi definita.

1.8. $\mathbb{Q}_+$, $\mathbb{R}_+$ (cioè i razionali positivi, i reali positivi) sono gruppi commutativi rispetto alla moltiplicazione.

1.9. Le n radici n-esime (complesse) di 1 formano un gruppo commutativo (finito) rispetto alla moltiplicazione definita in $\mathbb{C}$.

1.10. L'insieme delle funzioni biiettive da A su A è un gruppo *non* commutativo rispetto al prodotto di composizione $f \circ g$ (l'elemento neutro è la funzione identità e l'inverso di f è la funzione inversa f^{-1}).

1.11. In Geometria, esempi di gruppi commutativi sono dati dall'insieme delle *traslazioni* del piano in sé e dall'insieme delle *rotazioni* del piano in sé col centro in O (il prodotto di due traslazioni, o di due rotazioni, si definisce come l'applica-

zione della seconda operazione al risultato ottenuto applicando la prima); mentre *non* è *commutativo* il gruppo degli spostamenti rigidi del piano in sé (rotazioni + traslazioni).

1.12. Le matrici simili (cioè con lo stesso numero di righe e lo stesso numero di colonne) formano un gruppo commutativo rispetto all'addizione definita sulle matrici. Le matrici quadrate con determinante non nullo formano un gruppo *non* commutativo rispetto al prodotto (righe per colonne).

Osserviamo ora che gli elementi di un corpo formano gruppo rispetto alla somma; rispetto al prodotto essi formano gruppo se si esclude l'elemento neutro della somma (lo zero) per il quale non si può definire l'inverso. Infine gli elementi di uno spazio lineare (cioè i vettori) formano gruppo abeliano rispetto alla somma.

1.4 Prodotto scalare in $\mathbb{R}^n$

Ogni allievo sa che lo spazio fisico $\mathbb{R}^3$ ha una struttura assai più ricca di quella che appare dalla definizione precedente. Infatti non solo si possono eseguire operazioni come somma di vettori e prodotto di vettori per scalari, ma si può, per esempio, proiettare un vettore sopra un altro o parlare di lunghezza di un certo vettore, operazioni cui sono legati i concetti fisici fondamentali di lavoro e di energia. Queste possibilità non conseguono dalla struttura algebrica di questo spazio, ma da una ulteriore struttura, comune agli spazi euclidei (e non solo ad essi, ma certo non a tutti gli spazi lineari) che ora illustriamo.

Definizione 1.7 - *Siano* $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, $\mathbf{y} = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$. *Si dice* **prodotto scalare** *o* **prodotto interno** *di* $\mathbf{x}$ *per* $\mathbf{y}$ *il numero reale, indicato con* $< \mathbf{x}, \mathbf{y} >$, (si legga: $\mathbf{x}$ scalare $\mathbf{y}$ o $\mathbf{x}$ interno $\mathbf{y}$) *così definito*:

$$< \mathbf{x}, \mathbf{y} > := \sum_{i=1}^{n} x_i y_i \ . \tag{1.5}$$

L'applicazione: $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ definita dalla (1.5) possiede le seguenti proprietà, di verifica immediata: $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ e $\forall \lambda \in \mathbb{R}$ abbiamo

$\mathcal{S}$ 1. $< \mathbf{x}, \mathbf{y} > = < \mathbf{y}, \mathbf{x} >$
2. $< \mathbf{x} + \mathbf{y}, \mathbf{z} > = < \mathbf{x}, \mathbf{z} > + < \mathbf{y}, \mathbf{z} >$
3. $< \lambda \mathbf{x}, \mathbf{y} > = \lambda < \mathbf{x}, \mathbf{y} >$
4. $< \mathbf{x}, \mathbf{x} > \geq 0$; $< \mathbf{x}, \mathbf{x} > = 0 \Longleftrightarrow \mathbf{x} = \mathbf{0}$.

Dalle $\mathcal{S}$ 1. 2. 3. si ricava anche:

$$< \mathbf{x}, \lambda \mathbf{y} > = \lambda < \mathbf{x}, \mathbf{y} > \quad \text{e} \quad < \mathbf{x}, \mathbf{y} + \mathbf{z} > = < \mathbf{x}, \mathbf{y} > + < \mathbf{x}, \mathbf{z} > \ .$$

***Esempio* 1.13** - Se $\mathbf{F} = (F_1, F_2, F_3)$ è un vettore di $\mathbb{R}^3$ interpretabile come "forza" e $\mathbf{s} = (s_1, s_2, s_3)$ è un vettore "spostamento" il *prodotto* scalare $< \mathbf{F}, \mathbf{s} >$ rappresenta il lavoro compiuto da $\mathbf{F}$ dovuto allo spostamento $\mathbf{s}$ del suo punto di applicazione.

Le proprietà ($\mathcal{S}$) si ritengono distintive del prodotto scalare; cioè, se, in uno spazio lineare qualsiasi X (sul campo reale) è possibile definire un'applicazione: $X \times X \to \mathbb{R}$ dotata delle proprietà ($\mathcal{S}$), si dirà che in X è stato definito un prodotto scalare e lo *spazio* si dirà **pre-hilbertiano.**

Definizione 1.8 - *Sia* $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$. *Si chiama* **norma** *di* $\mathbf{x}$ *(o anche modulo, lunghezza, intensità) il numero reale non negativo, indicato con* $\|\mathbf{x}\|$, *così definito*

$$\|\mathbf{x}\| := \sqrt{< \mathbf{x}, \mathbf{x} >} = \sqrt{\sum_{i=1}^{n} x_i^2} \tag{1.6}$$

(se $\mathbf{x} \in \mathbb{R}^3$, la norma ora definita coincide con l'ordinaria nozione di lunghezza di un vettore). L'applicazione: $\mathbb{R}^n \to \mathbb{R}$, definita dalla (1.6), possiede le seguenti proprietà: $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\forall \lambda \in \mathbb{R}$, abbiamo

$\mathcal{N}$ 1. $\|\mathbf{x}\| \geq 0$; $\|\mathbf{x}\| = 0 \iff \mathbf{x} = \mathbf{0}$
2. $\|\lambda \mathbf{x}\| = |\lambda| \, \|\mathbf{x}\|$
3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (disuguaglianza triangolare)

Le $\mathcal{N}$1., 2. sono ovvie. La $\mathcal{N}$3. verrà dimostrata tra un momento. Osserviamo che, utilizzando la $\mathcal{N}$3., si ha

$$\|\mathbf{x}\| = \|(\mathbf{x} - \mathbf{y}) + \mathbf{y}\| \leq \|\mathbf{x} - \mathbf{y}\| + \|\mathbf{y}\|$$
$$\|\mathbf{y}\| = \|(\mathbf{y} - \mathbf{x}) + \mathbf{x}\| \leq \|\mathbf{y} - \mathbf{x}\| + \|\mathbf{x}\| = \|\mathbf{x} - \mathbf{y}\| + \|\mathbf{x}\|$$

e perciò risulta:

$$- \|\mathbf{x} - \mathbf{y}\| \leq \|\mathbf{x}\| - \|\mathbf{y}\| \leq \|\mathbf{x} - \mathbf{y}\| \, ,$$

cioè,

4. $\big| \, \|\mathbf{x}\| - \|\mathbf{y}\| \, \big| \leq \|\mathbf{x} - \mathbf{y}\|$.

Se $n = 1$ le $\mathcal{N}$3., 4. non sono altro che le (2.3.7) e (2.3.9). Per dimostrare la $\mathcal{N}$ 3. conviene premettere il seguente teorema.

■ **Teorema 1.1** - (disuguaglianza di Cauchy-Schwarz). *Siano* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$; *allora*

$$| < \mathbf{x}, \mathbf{y} > | \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\| \, . \tag{1.7}$$

Dimostrazione - Possiamo certamente supporre $\mathbf{x}, \mathbf{y} \neq \mathbf{0}$; diversamente, la (1.7) è ovvia. Sia λ reale qualsiasi; abbiamo

$$0 \le \|\mathbf{x} + \lambda\mathbf{y}\|^2 = <\mathbf{x} + \lambda\mathbf{y}, \mathbf{x} + \lambda\mathbf{y}> = <\mathbf{x},\mathbf{x}> + 2\lambda <\mathbf{x},\mathbf{y}> + \lambda^2 <\mathbf{y},\mathbf{y}> =$$
$$= \|\mathbf{x}\|^2 + 2\lambda <\mathbf{x},\mathbf{y}> + \lambda^2\|\mathbf{y}\|^2 .$$

Prendendo ora per λ il numero $- <\mathbf{x},\mathbf{y}> /\|\mathbf{y}\|^2$ la precedente espressione diventa:

$$\|\mathbf{x}\|^2 - \frac{2 <\mathbf{x},\mathbf{y}>^2}{\|\mathbf{y}\|^2} + \frac{<\mathbf{x},\mathbf{y}>^2}{\|\mathbf{y}\|^2} = \|\mathbf{x}\|^2 - \frac{<\mathbf{x},\mathbf{y}>^2}{\|\mathbf{y}\|^2} .$$

Essendo essa ≥ 0 troviamo precisamente la (1.7). □

Si osservi che, a parte il caso ovvio che uno dei due fattori $\mathbf{x}$ o $\mathbf{y}$ sia nullo, nella (1.7) vale il segno $=$ se e solo se $\mathbf{x} = \alpha\mathbf{y}$ oppure $\mathbf{y} = \beta\mathbf{x}$ (per qualche numero reale α o β).

Dalla (1.7) segue allora che, essendo il rapporto $<\mathbf{x},\mathbf{y}> /\|\mathbf{x}\| \cdot \|\mathbf{y}\|$ compreso tra -1 e $+1$, esiste un solo angolo θ, $0 \le \theta \le \pi$, tale che

$$\cos\theta = \frac{<\mathbf{x},\mathbf{y}>}{\|\mathbf{x}\| \cdot \|\mathbf{y}\|} . \tag{1.8}$$

Questo si chiama l'angolo dei due vettori. Il prodotto scalare $<\mathbf{x},\mathbf{y}>$ dei due vettori è così uguale al prodotto dei moduli $\|\mathbf{x}\| \cdot \|\mathbf{y}\|$ per il coseno dell'angolo compreso, in accordo con la consueta definizione data in Fisica.

Due vettori (non nulli) $\mathbf{x}, \mathbf{y}$ diconsi *ortogonali* se $<\mathbf{x},\mathbf{y}> = 0$.

Se calcoliamo allora

$$\|\mathbf{x} + \mathbf{y}\|^2 = <\mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y}> = <\mathbf{x},\mathbf{x}> + <\mathbf{x},\mathbf{y}> + + <\mathbf{y},\mathbf{x}> + <\mathbf{y},\mathbf{y}> \tag{1.9}$$

e supponiamo $\mathbf{x}$ e $\mathbf{y}$ ortogonali, abbiamo

$$\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 \tag{1.10}$$

cioè il classico *teorema di Pitagora*.

Sempre dalla (1.9), non supponendo i vettori ortogonali, ma usando la (1.7), otteniamo

$$\|\mathbf{x} + \mathbf{y}\|^2 \le \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\,\|\mathbf{y}\| + \|\mathbf{y}\|^2 = (\|\mathbf{x}\| + \|\mathbf{y}\|)^2$$

da cui si ricava la $\mathcal{N}3$.

La base canonica $\mathbf{e}^1, \mathbf{e}^2, \ldots, \mathbf{e}^n$ di $\mathbb{R}^n$ ha l'importante proprietà espressa dalla formula

$$<\mathbf{e}^j, \mathbf{e}^k> = \delta_{jk} \qquad j,k = 1,2,\ldots,n$$

dove δ_{jk} è il simbolo di Kronecker definito da:

$$\delta_{jk} := \begin{cases} 0 & \text{se } j \neq k \\ 1 & \text{se } j = k. \end{cases} \tag{1.11}$$

In altri termini, ogni $\mathbf{e}^j$ ha norma unitaria (si dice che è un *versore*) ed è ortogonale a tutti gli altri. Per questo la base si dice *ortonormale*.

Osserviamo ancora che, se $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, si ha

$$< \mathbf{x}, \mathbf{e}^j >= x_j$$

e quindi vale la formula di decomposizione

$$\mathbf{x} = \sum_{j=1}^{n} < \mathbf{x}, \mathbf{e}^j > \mathbf{e}^j \ . \tag{1.12}$$

Ogni termine $< \mathbf{x}, \mathbf{e}^j > \mathbf{e}^j$ è il vettore *proiezione* di $\mathbf{x}$ nella direzione $\mathbf{e}^j$.

La formula (1.12) vale con ogni base ortonormale, come l'allievo potrà facilmente verificare.

Osserviamo infine che le proprietà ($\mathcal{N}$) si ritengono distintive della norma; nel senso che, se in uno spazio lineare qualsiasi X (sul campo reale) è possibile definire un'applicazione: $X \to \mathbb{R}$ dotata delle proprietà ($\mathcal{N}$), si dice che per gli elementi di X è stata definita una norma e lo *spazio* si dirà **normato**. Per quanto visto, grazie alla disuguaglianza di Schwarz, ogni spazio dotato di prodotto scalare è automaticamente normato adottando per norma di $\mathbf{x}$ l'espressione $\sqrt{< \mathbf{x}, \mathbf{x} >}$; non è vero il viceversa.

Tramite la norma, possiamo ora definire la *distanza* tra due vettori.

Definizione 1.9 - *Siano* $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, $\mathbf{y} = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$. *Si dice* **distanza** *di* $\mathbf{x}$ *da* $\mathbf{y}$ *il numero reale non negativo, definito da:*

$$d(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \ . \tag{1.13}$$

L'applicazione: $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, definita dalla (1.13), possiede le seguenti proprietà: $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$, abbiamo

$\mathcal{D}$ 1. $d(\mathbf{x}, \mathbf{y}) \geq 0$; $d(\mathbf{x}, \mathbf{y}) = 0 \iff \mathbf{x} = \mathbf{y}$

2. $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$

3. $d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$ (disuguaglianza triangolare)

Le $\mathcal{D}$ 1., 2. sono ovvie. La $\mathcal{D}$ 3. discende dalla $\mathcal{N}$3.:

$$\begin{aligned} d(\mathbf{x}, \mathbf{y}) &= \|\mathbf{x} - \mathbf{y}\| = \|(\mathbf{x} - \mathbf{z}) + (\mathbf{z} - \mathbf{y})\| \leq \|\mathbf{x} - \mathbf{z}\| + \|\mathbf{z} - \mathbf{y}\| = \\ &= d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y}) \ . \end{aligned}$$

La distanza definita dalla (1.13) viene detta *pitagorica* o *euclidea*. Essa coincide in $\mathbb{R}^3$ con l'ordinaria nozione di distanza.

***Esempio* 1.14** - La distanza di due vettori qualsiasi della base canonica è $\sqrt{2}$:

$$d(\mathbf{e}^j, \mathbf{e}^k) = \sqrt{2} \quad (j \neq k; j, k = 1, 2, \ldots, n) .$$

Come meglio vedremo in seguito, in uno stesso spazio X si possono introdurre diverse definizioni di distanza, ma i requisiti $(\mathcal{D})$ sono da ritenersi essenziali perché una certa espressione possa essere adottata come distanza tra due punti. Si noti anche che, per la definizione di distanza (come applicazione: $X \times X \to \mathbb{R}$ dotata delle proprietà $(\mathcal{D})$) *non* si richiede che X abbia la struttura di spazio lineare. Un insieme qualsiasi X si dirà **spazio metrico** quando in esso sia stata introdotta una funzione distanza. Se X è uno spazio lineare, si dirà allora *spazio lineare metrico*. Per quanto visto, ogni spazio normato è anche metrico, adottando per distanza quella indotta dalla norma: $d(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\|$.

Esercizio - Verificare che in $\mathbb{R}^2$ le funzioni:

$$d_1(\mathbf{x}, \mathbf{y}) := |x_1 - y_1| + |x_2 - y_2|$$
$$d_2(\mathbf{x}, \mathbf{y}) := \max\{|x_1 - y_1|, |x_2 - y_2|\}$$

sono entrambe accettabili distanze.

Definizione 1.10 - *Sia* $\mathbf{x} \in \mathbb{R}^n$ *e* r *reale* > 0. *Si dice* **sfera**, *di centro* $\mathbf{x}$ *e raggio* r, *l'insieme*

$$S(\mathbf{x}, r) := \{\mathbf{y} \in \mathbb{R}^n : d(\mathbf{x}, \mathbf{y}) = r\} \qquad (1.14) .$$

Si dice **intorno sferico** *di* $\mathbf{x}$ *(o* **palla** *di centro* $\mathbf{x}$*) di raggio* r *l'insieme*

$$B(\mathbf{x}, r) := \{\mathbf{y} \in \mathbb{R}^n : d(\mathbf{x}, \mathbf{y}) < r\} . \qquad (1.15)$$

In $\mathbb{R}$ gli intorni sferici di x non sono altro che gli intervalli $(x - r, x + r)$.
In $\mathbb{R}^2$ gli intorni sferici di $\mathbf{x}$ sono cerchi centrati in $\mathbf{x}$, *privi della circonferenza*.
In $\mathbb{R}^3$ sono sfere piene, centrate in $\mathbf{x}$, *prive del bordo*.

a) intorno "sferico di x" in R *b)* intorno "sferico" di x in R² *c)* intorno "sferico" di x in R³

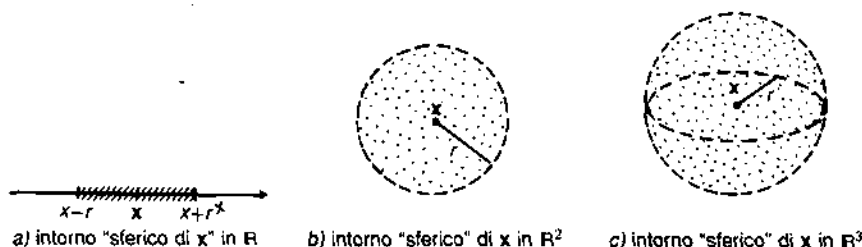



Fig. 3.5

Si può definire, più in generale, **intorno** di x un qualunque insieme che contenga un intorno sferico di $\mathbf{x}$. Ma per i nostri scopi saranno sufficienti gli intorni sferici.

Quando, per gli elementi di un insieme qualsiasi X, viene introdotta la nozione di *intorno*, si dice che è stata definita una *topologia* di X e l'insieme X si dice **spazio topologico** (*spazio lineare topologico* se X è uno spazio lineare). Se X è uno spazio metrico, dalla metrica di X discende subito la topologia, come sopra si è visto nel caso di $\mathbb{R}^n$. Ma gli intorni possono anche essere introdotti direttamente, mediante un sistema di assiomi, senza cioè farli derivare da una preesistente metrica. Un possibile sistema di assiomi è il seguente (altre formulazioni sono possibili, ma su questo argomento non intendiamo dilungarci):

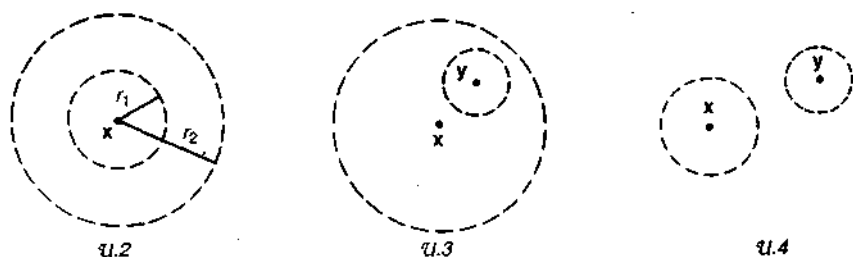



Fig. 3.6

$\mathcal{U}$ 1. Ogni punto x ha almeno un intorno $B(\mathbf{x})$; ogni intorno di $\mathbf{x}$ contiene $\mathbf{x}$.

2. L'intersezione $B_1(\mathbf{x}) \cap B_2(\mathbf{x})$ di due intorni di $\mathbf{x}$ contiene almeno un intorno di $\mathbf{x}$.

3. Se $\mathbf{y}$ appartiene a un intorno $B(\mathbf{x})$ di $\mathbf{x}$, esiste almeno un intorno $B(\mathbf{y})$ di $\mathbf{y}$ contenuto in $B(\mathbf{x})$.

4. Se $\mathbf{x} \neq \mathbf{y}$, esistono almeno un intorno $B(\mathbf{x})$ di $\mathbf{x}$ e un intorno $B(\mathbf{y})$ di $\mathbf{y}$ privi di punti comuni (proprietà di separazione di Hausdorff):

$$B(\mathbf{x}, r_1) \cap B(\mathbf{y}, r_2) = \emptyset .$$

Esercizio - Quali sono gli intorni "sferici" di $\mathbb{R}^2$ indotti dalle metriche d_1 e d_2 dell'esercizio precedente?

Le nozioni ora introdotte di prodotto scalare, norma, distanza, intorno, descrivono compiutamente la struttura di $\mathbb{R}^n$. Esse sono state introdotte in modo che dalla prima discendano via via le successive. L'allievo incontrerà, nel corso dei suoi studi, molti esempi di spazi lineari; ma non tutti avranno la ricchezza strutturale posseduta dagli spazi euclidei. Per ora limitiamoci a mostrare come le precedenti definizioni, date in uno spazio euclideo sul campo reale, possano estendersi agli spazi euclidei sul campo complesso.

1.5 Prodotto scalare in $\mathbb{C}^n$

Il *prodotto scalare* in $\mathbb{C}^n$ si definisce come l'applicazione: $\mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$ data da

$$< \mathbf{x},\mathbf{y} >:= \sum_{i=1}^{n} x_i \overline{y}_j \; . \tag{1.16}$$

Essa, possiede le seguenti proprietà: $\forall \mathbf{x},\mathbf{y},\mathbf{z} \in \mathbb{C}^n$, e $\forall \lambda \in \mathbb{C}$ abbiamo:

$\mathcal{S}$ 1. $< \mathbf{x},\mathbf{y} >= \overline{< \mathbf{y},\mathbf{x} >}$

2. $< \mathbf{x}+\mathbf{y},\mathbf{z} >=< \mathbf{x},\mathbf{z} > + < \mathbf{y},\mathbf{z} >$

3. $< \lambda\mathbf{x},\mathbf{y} >= \lambda < \mathbf{x},\mathbf{y} >$

4. $< \mathbf{x},\mathbf{x} >$ reale ≥ 0 ; $< \mathbf{x},\mathbf{x} >= 0 \Longleftrightarrow \mathbf{x} = \mathbf{0}$

2'. $< \mathbf{x},\mathbf{y}+\mathbf{z} >=< \mathbf{x},\mathbf{y} > + < \mathbf{x},\mathbf{z} >$

3'. $< \mathbf{x},\lambda\mathbf{y} >= \overline{\lambda} < \mathbf{x},\mathbf{y} >$.

Le 2' e 3' si deducono dalle 1. 2. 3. Si noti la 1. (detta proprietà di simmetria hermitiana) che è la proprietà distintiva del prodotto scalare sul campo complesso rispetto al campo reale. Anche nel campo complesso vale la *disuguaglianza di Schwarz*; infatti, riprendendo la dimostrazione precedente, abbiamo:

$$\begin{aligned} 0 \leq &< \mathbf{x}+\lambda\mathbf{y} \; , \; \mathbf{x}+\lambda\mathbf{y} >= \\ &=< \mathbf{x},\mathbf{x} > +\lambda < \mathbf{y},\mathbf{x} > +\overline{\lambda} < \mathbf{x},\mathbf{y} > +\lambda\,\overline{\lambda} < \mathbf{y},\mathbf{y} > \; . \end{aligned}$$

Prendendo ora per λ il numero complesso $- < \mathbf{x},\mathbf{y} > / < \mathbf{y},\mathbf{y} >$ l'ultima espressione diventa: $<\mathbf{x},\mathbf{x}> -|<\mathbf{x},\mathbf{y}>|^2/ <\mathbf{y},\mathbf{y}>$; imponendo che risulti non negativa ricaviamo

$$| < \mathbf{x},\mathbf{y} > | \leq \sqrt{< \mathbf{x},\mathbf{x} >} \cdot \sqrt{< \mathbf{y},\mathbf{y} >} \; . \tag{1.17}$$

La *norma* di un vettore $\mathbf{x} \in \mathbb{C}^n$ è definita ancora grazie alla $\mathcal{S}$ 4. e alla (1.17):

$$\|\mathbf{x}\| = \sqrt{< \mathbf{x},\mathbf{x} >} = \sqrt{\sum_{i=1}^{n} |x_i|^2} \; . \tag{1.18}$$

Essa definisce un'applicazione: $\mathbb{C}^n \to \mathbb{R}$ che possiede le stesse proprietà $\mathcal{N}$ 1., 2., 3. precedentemente elencate (dove naturalmente $|\lambda|$ significa ora: modulo del numero complesso λ). Per dimostrare la $\mathcal{N}$ 3. basta calcolare

$$\begin{aligned} \|\mathbf{x}+\mathbf{y}\|^2 &=< \mathbf{x},\mathbf{x} > + < \mathbf{x},\mathbf{y} > + < \mathbf{y},\mathbf{x} > + < \mathbf{y},\mathbf{y} >= \\ &=< \mathbf{x},\mathbf{x} > +2\,\mathrm{Re} < \mathbf{x},\mathbf{y} > + < \mathbf{y},\mathbf{y} >\leq \end{aligned}$$

(ricordando che $\mathrm{Re}\, z \leq |z|$)

$$\leq \|\mathbf{x}\|^2 + 2|< \mathbf{x},\mathbf{y} >| + \|\mathbf{y}\|^2 \leq$$

(applicando la disuguaglianza di Schwarz (1.17))

$$\leq \|\mathbf{x}\|^2 + 2\|\mathbf{x}\| \cdot \|\mathbf{y}\| + \|\mathbf{y}\|^2 = (\|\mathbf{x}\| + \|\mathbf{y}\|)^2 .$$

La *distanza* di due punti $\mathbf{x}$ e $\mathbf{y}$ in $\mathbb{C}^n$ è data da:

$$d(\mathbf{x},\mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2} \tag{1.19}$$

L'applicazione: $\mathbb{C}^n \times \mathbb{C}^n \to \mathbb{R}$ data dalla (1.19) ha ancora le proprietà $\mathcal{D}$ 1. 2. 3. Infine, ancora con le (1.14) e (1.15) si definiscono le *sfere* e gli *intorni sferici* in $\mathbb{C}^n$.

In particolare si osservi che la metrica (e la topologia) di $\mathbb{C}$ sono le stesse utilizzate per $\mathbb{R}^2$: in altre parole $\mathbb{R}^2$ e $\mathbb{C}$ possono identificarsi come spazi topologici (ma non come spazi lineari).

Esercizi

1. Determinare la dimensione del sottospazio di $\mathbb{R}^4$ generato dai vettori

$$\mathbf{v}^1 = (1,0,1,1) \ , \ \mathbf{v}^2 = (0,1,1,1) \ , \ \mathbf{v}^3 = (1,1,1,1) \ , \ \mathbf{v}^4 = (1,-1,0,0) \ .$$

2. Determinare quale/i tra i seguenti insiemi sono sottospazi di $\mathbb{R}^3$:

a) $\{(x,y,z) \in \mathbb{R}^3 : y < 0\}$

b) $\{(x,y,z) \in \mathbb{R}^3 : 3x + y + 2z = 0\}$

c) $\{(x,y,z) \in \mathbb{R}^3 : x + z = 0\}$

d) $\{(x,y,z) \in \mathbb{R}^3 : z = x^2 + y^2\}$

e) $\{(x,y,z) \in \mathbb{R}^3 : 3x + y + 2z = 1\}$.

3. Dimostrare che, fissato $\mathbf{h} \in \mathbb{R}^n$, l'insieme $\{\mathbf{x} \in \mathbb{R}^n : < \mathbf{h}, \mathbf{x} >= 0\}$ è un sottospazio di $\mathbb{R}^n$. Determinarne la dimensione.

4. Sia $\mathbf{M}$ una matrice di ordine $m \times n$.
Definiamo una funzione $\mathbf{f}: \mathbb{R}^n \to \mathbb{R}^m$ ponendo per ogni $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{f}(\mathbf{x}) = \mathbf{Mx}$ (prodotto righe per colonne).
Interpretare per $\mathbf{y} \in \mathbb{R}^n$ il teorema di Rouché-Capelli sulla risoluzione del sistema $\mathbf{Mx} = \mathbf{y}$ di m equazioni lineari in n incognite in termini di iniettività e suriettività per la funzione $\mathbf{f}$.

5. Sia $\mathbf{M}$ una matrice $m \times n$; sia $\mathbf{f}(\mathbf{x}) = \mathbf{Mx}$ (prodotto righe per colonne).
Dimostrare che i seguenti insiemi:

$\ker \mathbf{f} := \{\mathbf{x} \in \mathbb{R}^n : \mathbf{f}(\mathbf{x}) = 0\}$

(ker sta per "Kernel" (= nucleo, in tedesco)).

$\operatorname{im} \mathbf{f} := \{\mathbf{y} \in \mathbb{R}^m : \exists\, \mathbf{x} \in \mathbb{R}^n \text{ tale che } \mathbf{f}(\mathbf{x}) = \mathbf{y}\}$

sono sottospazi lineari di $\mathbb{R}^n$ ed $\mathbb{R}^m$ rispettivamente.

6. Sia M l'insieme delle matrici di ordine $m \times n$ ad elementi reali. Verificare che rispetto alle solite operazioni di somma e di moltiplicazione per un numero reale, M è uno spazio lineare su $\mathbb{R}$.

Determinarne la dimensione.

7. Verificare che $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ si ha:

$$\|\mathbf{x}+\mathbf{y}\|^2 + \|\mathbf{x}-\mathbf{y}\|^2 = 2\|\mathbf{x}\|^2 + 2\|\mathbf{y}\|^2 \quad \text{(legge del parallelogramma)} \tag{1.20}$$

Per $n = 2$ interpretare geometricamente la (1.20) in riferimento alla figura 3.7.

8. Verificare che le due funzioni da $\mathbb{R}^n \to \mathbb{R}$ definite nel modo seguente:

a) $\mathbf{x} \mapsto \underset{i=1,\dots,n}{\mathrm{Max}} |x_i| \quad (\mathbf{x} = (x_1, \dots, x_n))$

b) $\mathbf{x} \mapsto \sum_{i=1}^{n} |x_i|$

verificano le proprietà $\mathcal{N}_1$, $\mathcal{N}_2$, $\mathcal{N}_3$ delle norme. Qual è l'espressione della distanza tra due punti, indotta da ciascuna delle due norme precedenti?

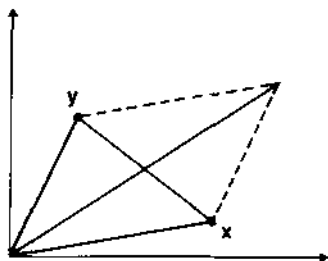



Fig. 3.7

9. Dimostrare che se $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ e

$$< \mathbf{x}+\mathbf{y} \ , \ \mathbf{x}-\mathbf{y} > = 0$$

allora $\|\mathbf{x}\| = \|\mathbf{y}\|$.

10. Definiamo in $\mathbb{R}$ la seguente nozione di *intorno*: $\forall x \in \mathbb{R}$ chiamiamo *intorno* di x ogni insieme del tipo

$$E_a := \{y \in \mathbb{R} : y > a \quad \text{dove} \quad a < x\} \ .$$

E_a

a x

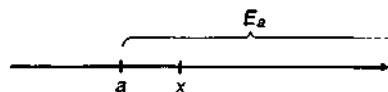



Fig. 3.8

Verificare che gli insiemi E_a soddisfano le proprietà $\mathcal{U}_1$, $\mathcal{U}_2$, $\mathcal{U}_3$ degli intorni ma non la proprietà di separazione $\mathcal{U}_4$.

11. Definiamo in $\mathbb{R}^2$ la seguente nozione di intorno: $\forall \mathbf{x} \in \mathbb{R}^2$ chiamiamo *intorno* di $\mathbf{x}$ ogni *disco (*) aperto con centro nell'origine* che contiene $\mathbf{x}$.

Verificare che gli intorni così definiti soddisfano le proprietà $\mathcal{U}_1$, $\mathcal{U}_2$, $\mathcal{U}_3$ ma non la proprietà di separazione $\mathcal{U}_4$.

2. ELEMENTI DI TOPOLOGIA IN $\mathbb{R}^n$

Esamineremo la topologia di $\mathbb{R}^n$ che abbiamo introdotto nel precedente paragrafo.

Ricordiamo che, parlando di intorno, intenderemo, salvo avviso contrario, gli intorni sferici.

2.1 Punti interni, esterni, di frontiera, di accumulazione, isolati

Sia dato un insieme $E \subseteq \mathbb{R}^n$; CE indicherà il complementare di E rispetto a tutto $\mathbb{R}^n$.

Definizione 2.1 - *Un punto* $\mathbf{x}$ *si dice* **interno** *ad* E *se esiste un suo intorno* $B(\mathbf{x}, r)$, $r > 0$, *contenuto in* E. *Si dice* **esterno** *ad* E *se è interno a* CE. *Si dice di* **frontiera** *per* E *se non è interno né esterno ad* E.

Prima di fare esempi osserviamo che:

– se $\mathbf{x}$ è interno ad E allora $\mathbf{x} \in E$;

– se $\mathbf{x}$ è esterno ad E allora $\mathbf{x} \notin E$;

– se $\mathbf{x}$ è di frontiera per E può essere $\mathbf{x} \in E$ oppure $\mathbf{x} \notin E$. In ogni caso *qualunque* intorno di $\mathbf{x}$ contiene sia punti di E sia punti di CE.

L'insieme dei punti interni di E si indica con $\hat{E}$; quello dei suoi punti di frontiera con ∂E. Evidentemente $\partial E = \partial CE$.

Esempi:

2.1. In $\mathbb{R}$, consideriamo i seguenti casi:

a) $E = (a, b)$. Ogni punto di E è *interno*: $E = E$; a e b sono di *frontiera*.

b) $E = [a, b)$. Ogni punto di E, tranne a, è *interno*; a e b sono di *frontiera*.

c) $E = \{x \in \mathbb{R} : x > a\}$. Ogni punto di E è *interno*; a è l'unico punto di *frontiera*.

d) $E = \mathbb{R}$. Ogni punto di E è *interno*. Non ci sono punti di *frontiera*.

(*) Si chiama *disco* una palla in dimensione 2.

e) $E = \mathbb{Q}$ (numeri razionali). Ogni punto di $\mathbb{R}$ è di *frontiera* per $\mathbb{Q}$: $\mathring{\mathbb{Q}} = \emptyset$, $\partial\,\mathbb{Q} = \mathbb{R}$. Si noti che $\mathbb{Q} \subset \partial\,\mathbb{Q}$.

2.2. In $\mathbb{R}^2$, consideriamo i seguenti casi:

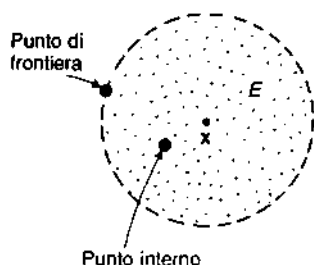



Fig. 3.9

a) $E = B(\mathbf{x}, r)$: E è un intorno sferico; tutti i suoi punti sono *interni*. I punti di frontiera, nessuno dei quali è in E, sono quelli della circonferenza di raggio r, centrata in $\mathbf{x}$;

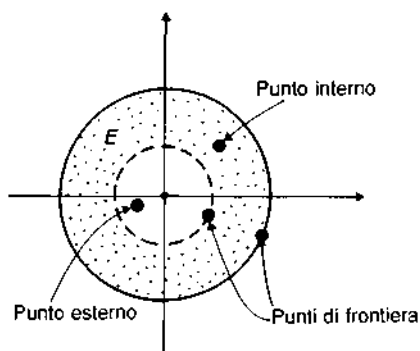



Fig. 3.10

b) $E = \{(x, y) \in \mathbb{R}^2; 1 < x^2 + y^2 \leq 4\}$. E descrive una corona circolare fra le due circonferenze di raggio 1 (che non è contenuta in E) e 2 (che è contenuta in E).
Punti interni ad E sono quelli della corona, escluse le circonferenze. La frontiera di E è costituita dall'unione delle due circonferenze;

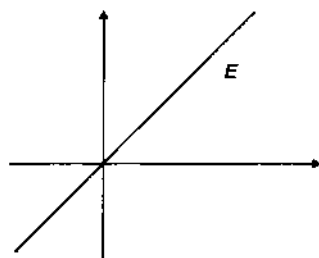



Fig. 3.11

c) $E = \{(x, y) \in \mathbb{R}^2; y = x\}$. E descrive la bisettrice del I° e III° quadrante. Non ci sono punti interni: $\mathring{E} = \emptyset$. E coincide con la propria frontiera: $E = \partial E$.

Della massima importanza è la definizione seguente.

Definizione 2.2 - *Un punto* $\mathbf{x}$ *si dice di* **accumulazione** *per E se in ogni intorno di* $\mathbf{x}$ *esiste un punto di E diverso da* $\mathbf{x}$.

Equivalentemente: ... *se in ogni intorno di* $\mathbf{x}$ *esistono infiniti punti di* E. (Dimostrare l'equivalenza delle due definizioni).

Un punto di accumulazione per E può appartenere o non appartenere a E. Evidentemente, tutti i punti interni di E sono di accumulazione per E. Se un punto di E non è di accumulazione (e allora deve essere di frontiera) si dirà **isolato**.

L'insieme dei punti di accumulazione di E si chiama il **derivato** di E, e si indica con E'. Se $E' = \emptyset$ (cioè non ci sono punti di accumulazione) E si dice **discreto**. È chiaro che tutti i punti di un insieme discreto sono isolati; ma non vale il viceversa, come subito vedremo. Se infine $E' = E$ (cioè l'insieme E è costituito da tutti e soli i suoi punti di accumulazione) E si dirà **perfetto**.

Esempi

2.3. In $\mathbb{R}$ consideriamo i seguenti casi:

a) $E = \mathbb{N}$ o $E = \mathbb{Z}$. I punti di E sono isolati in entrambi i casi; sia $\mathbb{N}$ che $\mathbb{Z}$ sono inoltre *discreti.*

b) $E = \left\{\frac{1}{n+1}; n \in \mathbb{N}\right\} = \left\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots\right\}$.

I punti di E sono isolati ma E *non* è discreto perché possiede un punto di accumulazione: lo zero.

c) $E = [a, b]$ è un insieme perfetto.

2.4. Riprendiamo gli insiemi dell'esempio 2.1:

a) I punti di E sono tutti di accumulazione; anche a e b lo sono: $E' = [a, b]$.

b) Come in a).

c) I punti di E sono tutti di accumulazione; anche a lo è.

d) $\mathbb{R}$ è un insieme perfetto.

e) Tutti i punti di $\mathbb{R}$ sono di accumulazione per $\mathbb{Q}$: $E' = \partial\,\mathbb{Q} = \mathbb{R}$.

2.5. In $\mathbb{R}^2$ consideriamo gli insiemi dell'esempio 2.2:

a) Ogni punto di E e della circonferenza (frontiera di E) è di accumulazione.

b) Ogni punto di E e di ∂E è di accumulazione: $E' = E \cup \partial E$.

c) Tutti e soli i punti di E sono di accumulazione: $E' = E$.

2.6. Ogni insieme $E \subset \mathbb{R}^n$ con un numero finito di elementi è discreto.

2.2 Insiemi aperti, chiusi, limitati

Definizione 2.3 - *Un insieme* $E \subseteq \mathbb{R}^n$ *si dice* **aperto** *se ogni* $\mathbf{x} \in E$ *è punto interno, cioè, se* E *coincide con* $\mathring{E}$; E *si dice* **chiuso** *se* CE *è aperto.*

Tra gli insiemi aperti si annovera anche l'insieme $\emptyset$.

Esempi

2.7. $\mathbb{R}^n$ è aperto. Poiché anche $\emptyset$ è aperto e $\mathbb{R}^n = C\emptyset$ allora $\mathbb{R}^n$ è anche chiuso. Stesso discorso per $\emptyset$.

$\mathbb{R}^n$ e $\emptyset$ sono gli *unici* insiemi che sono sia *aperti* che *chiusi*.

2.8. Riprendiamo gli insiemi dell'esempio 2.1. Ora siamo in $\mathbb{R}$.

a) E è aperto.

b) E non è aperto; infatti $a \in E$ ma non è interno ad E. Neppure è chiuso; infatti $b \in CE$ ma non è interno a CE.

c) E è aperto.

d) Come abbiamo già visto $\mathbb{R}$ è aperto e chiuso.

e) $\mathbb{Q}$ non è aperto né chiuso; inoltre $\mathring{\mathbb{Q}} = \emptyset$.

2.9. Riprendiamo gli insiemi dell'esempio 2.2. Ora siamo in $\mathbb{R}^2$.

a) E è aperto.

b) E non è aperto in quanto i punti della circonferenza di raggio 2 sono tutti in E ma non sono interni ad E. Neppure E è chiuso perché a CE appartengono i punti della circonferenza di raggio 1 che non sono interni.

c) E è chiuso.

È interessante notare che se siamo in dimensione 1, $\mathbb{R}$ è aperto (oltre che chiuso) ma se siamo in dimensione $n > 1$ ogni retta è un sottoinsieme chiuso di $\mathbb{R}^n$.

2.10. Un piano in $\mathbb{R}^3$ è un insieme chiuso.

2.11. Ogni insieme discreto è chiuso. In particolare l'insieme costituito da un solo punto è chiuso.

2.12. Per ogni $E \subseteq \mathbb{R}^n$, ∂E è un insieme chiuso; infatti il complementare di ∂E è un aperto, essendo formato dai punti che sono o interni ad E o interni a CE.

Passiamo ad alcune proprietà caratteristiche degli insiemi aperti e di quelli chiusi.
Sia $\Im$ una famiglia di insiemi. Indichiamo con $\bigcup_{A\in\Im} A$ (o semplicemente $\cup\Im$) e con $\bigcap_{A\in\Im} A$ (o anche $\cap\Im$) l'unione, e, rispettivamente, l'intersezione, di tutti gli insiemi di questa famiglia.

▪ Teorema 2.1 -
i) *se $\Im$ è una qualsiasi famiglia di aperti, $\cup\Im$ è pure un aperto*
ii) *se $\Im$ è una famiglia finita di aperti, $\cap\Im$ è pure un aperto.*

Dimostrazione - i) Se $\mathbf{x} \in \cup\Im$, allora $\mathbf{x}$ appartiene ad almeno un aperto A di $\Im$; pertanto esiste un intorno di $\mathbf{x}$ contenuto in A e perciò contenuto in $\cup\Im$. ii) Basterà provare che

$A \cap B$ è aperto se A e B sono aperti. Infatti, sia $\mathbf{x} \in A \cap B$; allora esiste un intorno $B(\mathbf{x}, r_1) \subset A$ e un altro $B(\mathbf{x}, r_2) \subset B$; ma allora l'intorno $B(\mathbf{x}, r)$, con $r = \min\{r_1, r_2\}$ sarà contenuto in $A \cap B$. □

Osservazione 2.1 - L'intersezione di una famiglia infinita di aperti può non essere un aperto, come mostra il seguente esempio in $\mathbb{R}^1$: l'intersezione di tutti gli intervalli $(-1/n, 1/n)$ $(n = 1, 2, \ldots)$ è $\{0\}$.

In maniera analoga si dimostra anche il seguente

■ **Teorema 2.2** -
i) *se $\Im$ è una famiglia qualsiasi di chiusi, $\cap\Im$ è pure un chiuso*
ii) *se $\Im$ è una famiglia finita di chiusi, $\cup\Im$ è chiuso.*

Per gli insiemi chiusi abbiamo due importanti caratterizzazioni:

■ **Teorema 2.3** - *Le tre condizioni seguenti sono equivalenti per un insieme $E \subseteq \mathbb{R}^n$:*

i) *E è chiuso:*

ii) $\partial E \subseteq E$

iii) ogni punto di accumulazione di E appartiene ad E: $E' \subseteq E$.

Dimostrazione - Faremo vedere che i) ⇒ ii), ii) ⇒ iii) ed infine iii) ⇒ i).
i) ⇒ ii) Sia E chiuso e $\mathbf{x} \in \partial E$. Allora o $\mathbf{x}$ è isolato (e in tal caso $\mathbf{x} \in E$) oppure $\mathbf{x}$ è di accumulazione per E. Poiché $\mathbf{x} \in \partial CE = \partial E$ e CE è aperto, $\mathbf{x} \notin CE$; dunque ancora $\mathbf{x} \in E$.
ii) ⇒ iii) Sia $\partial E \subseteq E$ e $\mathbf{x}$ di accumulazione per E. Poiché $\mathbf{x}$ non è isolato è: o interno, e allora $\mathbf{x} \in E$, o di frontiera, ed ancora $\mathbf{x} \in E$, essendo $\partial E \subseteq E$.
iii) ⇒ i) Supponiamo che ogni punto di accumulazione di E appartenga ad E. Sia $\mathbf{x} \in CE$; vogliamo dimostrare che $\mathbf{x}$ è interno a CE. Se non lo fosse *ogni* intorno di $\mathbf{x}$ conterrebbe punti di E; dunque $\mathbf{x}$ sarebbe di accumulazione per E e, poiché $\mathbf{x} \in CE$, $\mathbf{x} \notin E$, contro l'ipotesi. □

Definizione 2.4 - *Sia $E \subseteq \mathbb{R}^n$; si chiama* **chiusura** *di E e si indica con $\overline{E}$, l'insieme $E \cup \partial E$.*

Evidentemente, dal teorema 2.3, E è chiuso se e solo se $E = \overline{E}$.

Risulta quindi, per ogni insieme $E \subseteq \mathbb{R}^n$:

$$\mathring{E} \subseteq E \subseteq \overline{E} \ , \ \overline{E} = E \cup \partial E = \mathring{E} \cup \partial E \ , \ \partial E = \overline{E} \setminus \mathring{E} \ .$$

Sia ora A un sottoinsieme di E. Se risulta $\overline{A} = \overline{E}$, diremo che A è **denso** in E.

***Esempio* 2.13** - a) Abbiamo già osservato, per l'insieme $\mathbb{Q} \subset \mathbb{R}$, che risulta $\partial\,\mathbb{Q} = \mathbb{R}$; risulta perciò che $\overline{\mathbb{Q}} = \mathbb{R}$ e dunque $\mathbb{Q}$ è denso in $\mathbb{R}$; ciò era già stato

dichiarato in 2.2.4; lo studente osservi che la definizione di insieme denso in $\mathbb{R}$ data in 2.2.4 è in accordo con quella (più generale) riportata sopra.

b) $\mathbb{R} \setminus \mathbb{Q}$ è denso in $\mathbb{R}$.

c) I punti di $\mathbb{R}^n$ con coordinate razionali sono densi in $\mathbb{R}^n$.

Se un insieme di $\mathbb{R}^n$ ammette un punto di accumulazione, esso contiene necessariamente infiniti elementi. Il viceversa non è vero come mostrano gli insiemi discreti infiniti ($\mathbb{N}, \mathbb{Z}, \ldots$). Per chiarire meglio la questione, cominciamo col dare la definizione di insieme limitato.

Abbiamo già parlato (cfr. 1.3.1) di insiemi limitati in un insieme ordinato. Ma il concetto di limitatezza prescinde dall'ordinamento, e può facilmente essere introdotto in $\mathbb{R}^n$.

Definizione 2.5 - *Un insieme $E \subset \mathbb{R}^n$ si dirà* **limitato** *se esiste un intorno dell'origine $B(\mathbf{0}, r)$, che lo contiene.*

La precedente definizione, applicata al caso dell'insieme (ordinato) $\mathbb{R}$ coincide con quella introdotta in 1.3.1.

Osservazione 2.2 - Dato $E \subseteq \mathbb{R}^n$, possiamo considerare l'insieme D dei numeri reali non negativi così definito: $D := \{\|\mathbf{x} - \mathbf{y}\| : \mathbf{x}, \mathbf{y} \in E\}$; se questo insieme è limitato superiormente il suo estremo superiore lo chiameremo *diametro* di E e scriveremo:

$$\operatorname{diam}(E) := \sup\{\|\mathbf{x} - \mathbf{y}\| : \mathbf{x}, \mathbf{y} \in E\} .$$

Se $E \subset \mathbb{R}^2$ è un cerchio, $\operatorname{diam}(E)$ è il diametro che tutti conoscono; se è un quadrato, è la lunghezza della diagonale.

Allora si osserverà che un insieme $E \subseteq \mathbb{R}^n$ è limitato se e solo se ha diametro finito. In particolare, $\mathbb{R}^n$ non è limitato.

Il seguente teorema ha un ruolo fondamentale nell'Analisi degli spazi euclidei $\mathbb{R}^n$.

▪ **Teorema 2.4** - (di Bolzano-Weierstrass). *Ogni $E \subset \mathbb{R}^n$, limitato e infinito, possiede almeno un punto di accumulazione.*

Dimostrazione - È divisa in due parti; nella prima si individua un punto candidato ad essere di accumulazione per E; nella seconda si mostra che il punto prima individuato è effettivamente di accumulazione per E. Esporremo la dimostrazione per $n = 2$, ma non c'è alcuna difficoltà ad adattarla a qualunque n.

I^a parte - Poiché E è limitato, esisteranno numeri p, q, r, s tali che $E \subset T = [p, q] \times [r, s]$ (v. fig. 3.12). Se si divide $[p, q]$ in due intervalli uguali e lo stesso si fa per $[r, s]$, almeno uno dei 4 rettangoli così ottenuti dovrà contenere infiniti punti di E.

Chiamiamo $T_1 = [p_1, q_1] \times [r_1, s_1]$ un tale rettangolo e ripetiamo su T_1 l'operazione precedente (questo si chiama procedere per *dicotomia*). Otterremo un rettangolo $T_2 =$

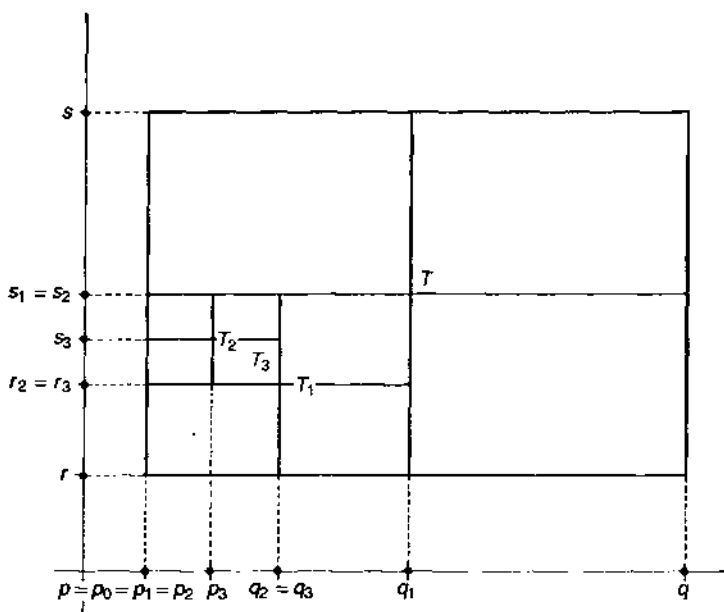



Fig. 3.12

$= [p_2, q_2] \times [r_2, s_2]$ contenente ancora infiniti punti di E e così via. Si costruisce così una successione di rettangoli $\{T_n\}$, ciascuno dei quali è contenuto nel precedente, e in ciascuno dei quali cadono infiniti punti di E.

Consideriamo ora i due insiemi

$$P = \{p_1, p_2, p_3, \ldots\} \quad \text{e} \quad Q = \{q_1, q_2, q_3, \ldots\} \, .$$

Dalla costruzione fatta risulta chiaramente che

$$p_1 \leq p_2 \leq p_3 \leq \ldots \quad \text{e} \quad q_1 \geq q_2 \geq q_3 \geq \ldots$$

e inoltre, per ogni intero h e k, sarà

$$p_h \leq p_n \leq q_n \leq q_k \quad \text{se } n \geq h \ , \ n \geq k$$

$$q_n - p_n = \frac{q - p}{2^n} \, .$$

Ciò significa che q_k è un maggiorante di P e perciò $\sup P \leq q_k$; ma dovendo valere questa disuguaglianza per ogni k, risulta che $\sup P$ è un minorante di Q e dunque: $\sup P \leq$ $\leq \inf Q$. Ma risulta anche (essendo $\inf Q \leq q_n$, $\sup P \geq p_n$ per ogni n)

$$\inf Q - \sup P \leq q_n - p_n = \frac{q - p}{2^n}$$

e perciò deve essere $\inf Q = \sup P$; chiamiamo x_1 questo valore comune.
Ripetendo le stesse considerazioni sugli insiemi $R = \{r_1, r_2, r_3, \ldots\}$ e $S = \{s_1, s_2, s_3, \ldots\}$ si trova che $\sup R = \inf S$; chiamiamo x_2 questo valore comune. Abbiamo così individuato il punto $\mathbf{x} = (x_1, x_2)$. Osserviamo che

$$\{\mathbf{x}\} = \bigcap_{k=1}^{\infty} T_k \,.$$

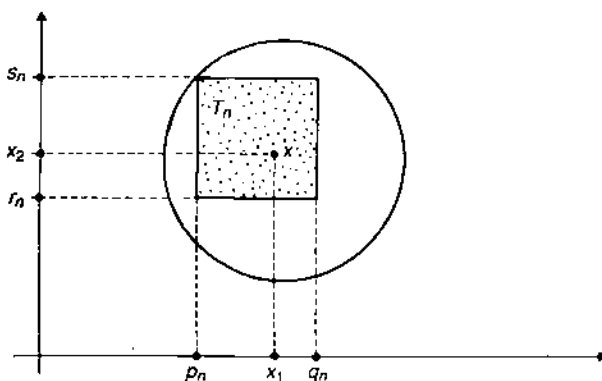



Fig. 3.13

II^a parte - Mostriamo che il punto $\mathbf{x}$ è di accumulazione per E. Sia perciò $B(\mathbf{x}, \rho)$ un intorno sferico di $\mathbf{x}$ (di raggio ρ), fissato arbitrariamente. È chiaro che per n sufficientemente grande il rettangolo T_n sarà contenuto in $B(\mathbf{x}, \rho)$: basterà, per esempio, prendere n tale che (v. figura 3.13)

$$q_n - p_n = \frac{q - p}{2^n} < \rho$$

$$s_n - r_n = \frac{s - r}{2^n} < \rho$$

Perciò $B(\mathbf{x}, \rho)$, contenendo T_n, contiene infiniti punti di E. Per l'arbitrarietà di ρ segue l'asserto. □

Osservazione 2.3 - Sugli insiemi **chiusi e limitati** in $\mathbb{R}$.

Sia $X \subset \mathbb{R}$ limitato; allora $\sup X$ e $\inf X$ sono finiti; se si tiene presente la definizione di estremo superiore (inferiore) si riconosce subito che $\sup X$ e $\inf X$ o sono punti isolati di X o sono punti di accumulazione per X (cfr. Esercizio 2.3); nel primo caso essi appartengono a X; nel secondo caso certamente apparterranno a X se X è chiuso; *perciò un insieme chiuso e limitato in* $\mathbb{R}$ *ammette massimo e minimo.*

2.3 La retta ampliata. Gli spazi $\dot{\mathbb{R}}^n$

In molte questioni, e soprattutto in vista della definizione di limite che daremo nel prossimo capitolo, conviene considerare, accanto alla retta $\mathbb{R}$, anche un suo ampliamento ottenuto aggiungendo ad $\mathbb{R}$ due nuovi punti (*non* numeri!) indicati coi simboli $+\infty$ e $-\infty$. Ciò che otteniamo è un nuovo insieme, che indicheremo con $\mathbb{R}^*$:

$$\mathbb{R}^* := \mathbb{R} \cup \{-\infty\} \cup \{+\infty\} .$$

Ad $\mathbb{R}^*$ immediatamente si estende l'ordinamento di $\mathbb{R}$ convenendo che, $\forall x \in \mathbb{R}$, sia

$$-\infty < x < +\infty$$

cosicché $\mathbb{R}^*$ è un insieme totalmente ordinato.

Un modello di $\mathbb{R}^*$ si può ottenere considerando le proiezioni dal centro C sulla retta r dei punti della semicirconferenza illustrata in figura 3.14.

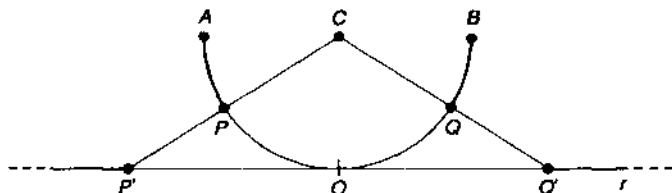



Fig. 3.14

Nella proiezione dal centro C ad A "corrisponde" $-\infty$, a B "corrisponde" $+\infty$.

Si osservi che ogni insieme in $\mathbb{R}^*$ risulta limitato, poiché esiste sempre un maggiorante $(+\infty)$ e un minorante $(-\infty)$. La proprietà $\mathcal{R}_4$ (di completezza di $\mathbb{R}$) si può riformulare così: *ogni insieme $A \subseteq \mathbb{R}^*$ non vuoto ammette estremo superiore ed estremo inferiore.* Il $\sup A$ sarà un numero reale se A è limitato superiormente in $\mathbb{R}$, altrimenti sarà $\sup A = +\infty$. Analogamente per l'estremo inferiore.

$\mathbb{R}^*$ può pure essere dotato di struttura topologica; basterà definire gli intorni di $+\infty$ e $-\infty$, come proiezione da C di archi PA e QB rispettivamente, come indicato in figura 3.14.

Chiameremo *intorno di* $+\infty$ qualunque semiretta del tipo $\{x \in \mathbb{R}^* : a < x \leq +\infty\}$ che denoteremo col simbolo $(a, +\infty]$. Analoga definizione e notazione ($[-\infty, b)$) per gli intorni di $-\infty$. Si noti che abbiamo scritto $(a, +\infty]$ e non $(a, +\infty)$, poiché vogliamo che ogni punto appartenga ad ogni suo intorno. La famiglia di intorni dei punti di $\mathbb{R}^*$ (quelli al finito precedentemente introdotti e quelli all'infinito definiti ora) verifica le proprietà $(\mathcal{U})$ di 1.4 e perciò

$\mathbb{R}^*$ è uno spazio topologico. Notiamo anche che, tra gli intorni sferici dell'origine, c'è anche l'intero spazio $\mathbb{R}^* = [-\infty, +\infty]$.

Pensando $\mathbb{R}$ e tutti i suoi sottoinsiemi come "immersi" nell'ampliamento $\mathbb{R}^*$, possiamo enunciare il teorema di Bolzano-Weierstrass nel seguente modo: *ogni sottoinsieme infinito di* $\mathbb{R}$ *ammette almeno un punto di accumulazione in* $\mathbb{R}^*$.

Ad esempio $\mathbb{N}$ ammette come (unico) punto di accumulazione in $\mathbb{R}^*$ il punto $+\infty$.

Osservazione 2.4 - Poiché $\mathbb{R}^*$ è uno spazio topologico, possiamo parlare di insiemi aperti, chiusi, limitati in $\mathbb{R}^*$ (e di punti interni, di frontiera, di accumulazione etc.) così come abbiamo definito questi concetti per gli insiemi di $\mathbb{R}$. Ma, essendo $\mathbb{R}^*$ ed $\mathbb{R}$ dotati ognuno della propria topologia, non deve meravigliare se insiemi aperti o chiusi o limitati in $\mathbb{R}$ non hanno la stessa proprietà in $\mathbb{R}^*$. Per esempio, si consideri l'insieme $\mathbb{N}$.

Pensando $\mathbb{N} \subset \mathbb{R}$ allora $\mathbb{N}$ è:
chiuso (non ha punti di accumulazione)
non limitato (non è contenuto in alcun intorno dell'origine).

Pensando $\mathbb{N} \subset \mathbb{R}^*$ allora $\mathbb{N}$ è:
non chiuso (poiché non contiene $+\infty$, che è punto di accumulazione)
limitato (poiché è contenuto nell'intorno $[-\infty, +\infty]$).

Purtroppo la struttura algebrica di $\mathbb{R}$ non può trasportarsi su $\mathbb{R}^*$, cioè non possiamo definire, sui simboli $+\infty$ e $-\infty$, le operazioni algebriche di somma e prodotto in modo che risultino verificate le usuali proprietà. Una aritmetizzazione parziale di questi simboli è possibile, come vedremo; ma la struttura di campo (né quella di spazio lineare) non è trasportabile su $\mathbb{R}^*$ (perciò $+\infty$ e $-\infty$ *non hanno diritto* ad essere chiamati numeri (né vettori)!). $\mathbb{R}^*$ è perciò uno *spazio topologico*, ma *non lineare*. Tuttavia, nonostante questo difetto, la considerazione di $\mathbb{R}^*$ risulterà molto utile per la definizione di limite.

In talune considerazioni risulta utile un diverso ampliamento di $\mathbb{R}$, che possiamo indicare con $\dot{\mathbb{R}}$, ottenuto da $\mathbb{R}$ con l'aggiunta di *un solo* punto, ∞:

$$\dot{\mathbb{R}} := \mathbb{R} \cup \{\infty\} .$$

Si definisce *intorno (sferico)* di ∞ il complementare di qualunque intervallo chiuso $[-a, a]$ (con $a > 0$); cosicché $\dot{\mathbb{R}}$ è dotato di struttura topologica. Ancora, tra gli intorni sferici di 0 possiamo includere l'intero spazio $\dot{\mathbb{R}}$. Possiamo notare che $\dot{\mathbb{R}}$ può mettersi in corrispondenza biunivoca con i punti di una circonferenza nel modo indicato in figura 3.15.

Al punto N viene così a corrispondere il punto ∞. Ciò aiuta a capire come in $\dot{\mathbb{R}}$ non solo non sia trasportabile la struttura algebrica di $\mathbb{R}$, ma neanche l'ordinamento $(*)$.

$(*)$ Lo studente dovrebbe già conoscere questo insieme, noto in Geometria col nome di *retta proiettiva*.

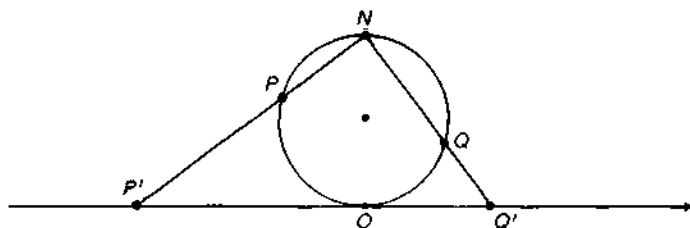



Fig. 3.15

Analogamente si può definire lo spazio topologico $\dot{\mathbb{R}}^n$, come l'insieme ottenuto da $\mathbb{R}^n$ con l'aggiunta di un punto, $\{\infty\}$, e reso topologico con la definizione seguente di intorno (sferico) di ∞: *il complementare di qualunque palla chiusa di* $\mathbb{R}^n$, *con centro in* $\mathbf{0}$. La collezione degli intorni $B(\mathbf{x})$ e $B(\infty)$ così definita verifica le proprietà ($\mathcal{U}$) di 1.4 e pertanto $\dot{\mathbb{R}}^n$ è uno spazio topologico (ma non lineare topologico). Pensando $\mathbb{R}^n$ come immerso nell'ampliamento $\dot{\mathbb{R}}^n$, il teorema di Bolzano-Weierstrass diventa: *ogni insieme infinito di* $\mathbb{R}^n$ *ammette almeno un punto di accumulazione in* $\dot{\mathbb{R}}^n$.

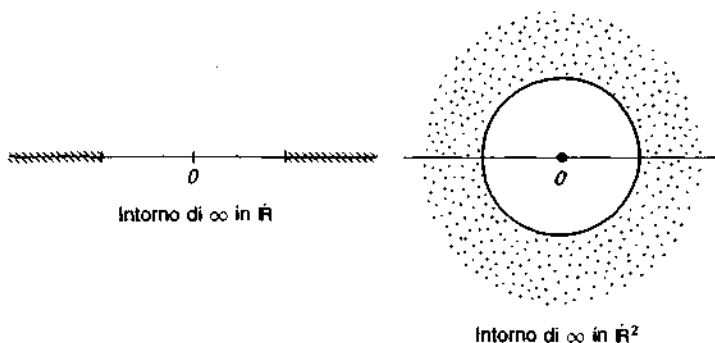



Fig. 3.16

2.4 Insiemi compatti

Sia $E \subseteq \mathbb{R}^n$. Sia $\mathfrak{I}$ una famiglia di aperti di $\mathbb{R}^n$ tale che:

$$\cup \mathfrak{I} \supseteq E.$$

Una tale famiglia viene detta una *copertura* di E.
Se alla copertura $\mathfrak{I}$ si aggiungono altri insiemi, la nuova famiglia è ancora una copertura, ovviamente; non così se si tolgono degli insiemi.

Definizione 2.6 - *Un insieme* $E \subseteq \mathbb{R}^n$ *si dirà* **compatto** *se da* **ogni** *sua copertura è possibile estrarre una famiglia* **finita** *di aperti che sia ancora una copertura.*

***Esempio* 2.14** - Sia $E = \mathbb{N} \subset \mathbb{R}$. Una copertura di $\mathbb{N}$ può ottenersi dall'unione degli aperti del tipo $(n - \varepsilon, n + \varepsilon)$ con $n \in \mathbb{N}$ e $0 < \varepsilon < 1$. $\mathbb{N}$ non è compatto; infatti se dalla precedente copertura togliamo anche un solo aperto non abbiamo più una copertura.

È pure facile vedere che l'intero spazio $\mathbb{R}^n$ non è compatto, mentre ogni insieme finito è compatto.

Il seguente teorema dà una caratterizzazione degli insiemi compatti in $\mathbb{R}^n$.

■ **Teorema 2.5** - (di Heine-Borel). *Sia $E \subset \mathbb{R}^n$; allora:*

E compatto $\iff$ *E chiuso e limitato* .

Ogni intervallo $[a, b]$, $a, b \in \mathbb{R}$, è compatto in $\mathbb{R}$.
La chiusura di ogni intorno sferico di un punto $\mathbf{x} \in \mathbb{R}^n$ è un compatto. Evidentemente, l'unione di un numero finito di compatti è un compatto.

Dimostrazione - Facciamo vedere che se E è compatto allora è chiuso e limitato. Ci limiteremo al caso $E \subset \mathbb{R}^2$. La dimostrazione si estende senza difficoltà a dimensioni qualunque.

E è limitato: infatti la famiglia di cerchi $\{B(\mathbf{0}, k)\}_{k \in \mathbb{N}}$ è una copertura di $\mathbb{R}^2$ e quindi anche di E. Essendo E compatto esiste una sottofamiglia finita che è ancora una copertura. Ma allora E è contenuto nel cerchio più grande di questa sottofamiglia e dunque è limitato.

E è chiuso: sia $\mathbf{x} \in CE$; per ogni $\mathbf{y} \in E$ costruiamo un intorno $U_\mathbf{y}$ di $\mathbf{x}$ e un intorno $V_\mathbf{y}$ di $\mathbf{y}$ tali che $U_\mathbf{y} \cap V_\mathbf{y} = \emptyset$. La famiglia $\{V_\mathbf{y}\}_{\mathbf{y} \in E}$ è una copertura di E dalla quale si può estrarre una sottocopertura finita $V_{\mathbf{y}_1}, \ldots, V_{\mathbf{y}_N}$. Sia $U = U_{\mathbf{y}_1} \cap U_{\mathbf{y}_2} \cap \ldots \cap U_{\mathbf{y}_N}$; allora $U \cap E = \emptyset$ e U è un intorno di $\mathbf{x}$. Ne segue che CE è aperto.

Sia ora E chiuso e limitato. Per far vedere che è compatto procediamo in 2 passi:

i) ogni rettangolo $E = [a, b] \times [c, d] \subseteq \mathbb{R}^2$ è compatto

ii) ogni sottoinsieme chiuso di un compatto è compatto.

Dim. di i): procediamo per assurdo e supponiamo che $\mathfrak{F}$ sia una copertura di E dalla quale *non* si possa estrarre una sottocopertura finita. Procediamo per dicotomia come nel Teorema di Bolzano-Weierstrass costruendo una successione di rettangoli T_k con le seguenti proprietà:

1) $T_{k+1} \subset T_k \quad \forall k \geq 1$

2) area $T_k = \dfrac{1}{4^k}(b - a)(d - c)$

3) *nessuna* sottofamiglia finita di $\mathfrak{F}$ è copertura di T_k.

Sia ora $\mathbf{z} = \bigcap_{k=1}^{\infty} T_k$. Poiché $\mathbf{z} \in E$ esiste un aperto $U \in \mathfrak{F}$ tale che $\mathbf{z} \in U$. Sia $B(\mathbf{z}, r)$ un intorno di $\mathbf{z}$ contenuto in U. Per k abbastanza grande si avrà $T_k \subset B(\mathbf{z}, r) \subset U$ che contraddice la 3).

Dim. di ii): sia E chiuso, $E \subseteq F$ con F compatto. Allora se $\Im$ è una copertura di E, $\Im \cup \complement E$ è una copertura di F. Essendo F compatto esiste una sottocopertura finita $U_1, \ldots, U_N, \complement E$. Allora $U_1, \ldots, U_N$ è una sottocopertura finita di E.

Per completare il teorema è sufficiente osservare che se E è limitato allora è contenuto in un rettangolo $[a,b] \times [c,d]$. Dunque E è chiuso ed è sottoinsieme di un compatto. Usando la ii) si conclude che E è compatto. □

Osservazione 2.4 - La definizione di compattezza si estende agli ampliamenti di $\mathbb{R}$ ed $\mathbb{R}^n$, tale e quale. Gli spazi $\mathbb{R}^*, \dot{\mathbb{R}}, \dot{\mathbb{R}}^n$ sono compatti rispetto alla propria topologia. Infatti, sia $\Im$ una copertura di $\dot{\mathbb{R}}$ (per esempio). Essa deve contenere un intorno di ∞, diciamo il complementare di $[-a, a]$. Con i rimanenti elementi di $\Im$ dobbiamo coprire questo intervallo che, essendo chiuso e limitato in $\mathbb{R}$, può coprirsi con un numero finito di aperti di $\Im$.

2.5 Insiemi connessi. Insiemi convessi

La nozione di connessione vuole rendere precisa l'idea intuitiva di insieme costituito di un solo pezzo; essa svolge una parte importante nello studio delle applicazioni continue oggetto del capitolo 5.

Siano A, B sottoinsiemi di $\mathbb{R}^n$. Essi si dicono **separati** *se:*

$$\overline{A} \cap B = \emptyset \quad \text{e} \quad A \cap \overline{B} = \emptyset \, .$$

Gli esempi in figura 3.17 illustrano il concetto per insiemi in $\mathbb{R}^2$.

***Esempio* 2.15** - Gli insiemi $A = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 > 0, x_1 > 0\}$ e $B = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 < 0, x_1 < 0\}$ sono separati.

A e *B* sono separati

A e *B*, pur essendo disgiunti, non sono separati

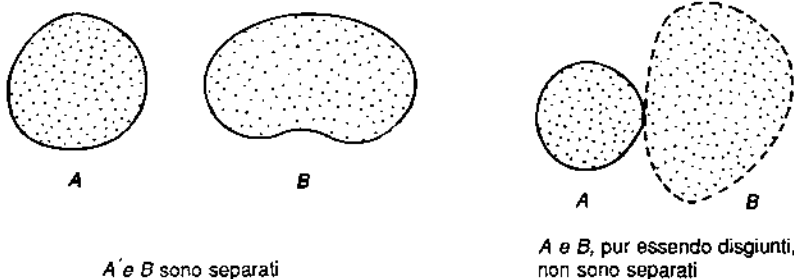



Fig. 3.17

In altre parole, A e B sono separati se e solo se non hanno elementi in comune, e inoltre nessun punto di A è di accumulazione per B e nessun punto di B è di accumulazione per A.

Definizione 2.7 - *Un insieme di* $\mathbb{R}^n$ *si dice* **connesso** *se non si può rappresentare come unione di due insiemi non vuoti e separati.*

***Esempio* 2.16** - L'insieme $C = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 x_2 > 0\}$ non è connesso, poiché risulta unione dei due insiemi A e B dell'esempio precedente. Più avanti mostreremo invece che l'insieme $D = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 x_2 \geq 0\}$ è connesso.

Un insieme aperto, non vuoto, connesso si dice **dominio** (*).

C'è una definizione di connessione un po' più restrittiva della precedente, ma più intuitiva e, per certi versi, più comoda. Ci servono le seguenti definizioni.

Se $\mathbf{x}$ *e* $\mathbf{y} \in \mathbb{R}^n$, *diremo* **retta** *passante per* $\mathbf{x}$ *e* $\mathbf{y}$ *l'insieme*

$$\{\lambda\mathbf{x} + \mu\mathbf{y} : \lambda, \mu \in \mathbb{R}, \lambda + \mu = 1\} ;$$

diremo **segmento** *(chiuso) congiungente* $\mathbf{x}$ *con* $\mathbf{y}$, *e lo indicheremo con* $[\mathbf{x}, \mathbf{y}]$, *l'insieme (combinazione lineare convessa di* $\mathbf{x}$ *e* $\mathbf{y}$*)*

$$\{\lambda\mathbf{x} + \mu\mathbf{y} : \lambda, \mu \in \mathbb{R}, \lambda \geq 0, \mu \geq 0, \lambda + \mu = 1\} .$$

L'allievo controllerà, per $n = 2$, $n = 3$, che queste definizioni si accordano con quanto egli già conosce dalla Geometria.

Diremo **spezzata poligonale** *congiungente* $\mathbf{x}$ *con* $\mathbf{y}$ *la riunione di un numero finito di segmenti del tipo:* $[\mathbf{x}, \mathbf{x}_1]$, $[\mathbf{x}_1, \mathbf{x}_2], \ldots, [\mathbf{x}_{n-1}, \mathbf{y}]$ *essendo* $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{n-1}$ *punti di* $\mathbb{R}^n$.

Dato un insieme $A \subseteq \mathbb{R}^n$, e preso $\mathbf{x} \in A$, indichiamo con $A[\mathbf{x}]$ il sottoinsieme di A costituito dai punti $\mathbf{y} \in A$ che si possono congiungere con $\mathbf{x}$ mediante una spezzata poligonale *tutta contenuta* in A.

Diremo che $A \subseteq \mathbb{R}^n$ *è* **connesso per segmenti** *se, preso un qualunque* $\mathbf{x} \in A$, *risulta* $A[\mathbf{x}] = A$.

***Esempio* 2.17** - L'insieme D dell'esempio precedente è connesso per segmenti. In $\mathbb{R}^2$, l'insieme dei punti con almeno una coordinata razionale è connesso per segmenti.

Abbiamo detto che la connessione per segmenti è più restrittiva della connessione; infatti si può dimostrare che, *se un insieme è connesso per segmenti, è anche connesso*; il viceversa non è vero; per esempio, una circonferenza in $\mathbb{R}^2$ è un insieme connesso, ma non connesso per segmenti.

Tuttavia, se limitiamo la nostra considerazione agli insiemi *aperti* di $\mathbb{R}^n$, le due definizioni si equivalgono: *ogni insieme aperto di* $\mathbb{R}^n$ *connesso, è connesso per segmenti.* Ma non riportiamo la dimostrazione di questi risultati.

Notiamo infine che esiste un'altra nozione di connessione, la *connessione per archi*, che ci sembra però prematuro presentare a questo punto.

(*) Ricordiamo però che la parola *dominio* è stata già usata con significato meno stretto in relazione all'insieme di definizione di una funzione.

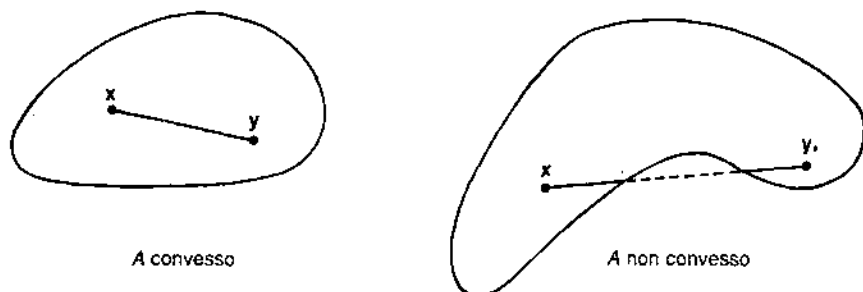



Fig. 3.18

Tra gli insiemi connessi hanno grande importanza quelli *convessi* così definiti.

Definizione 2.8 - *Un insieme* $A \subseteq \mathbb{R}^n$ *si dice* **convesso** *se, per ogni coppia di punti* $\mathbf{x}, \mathbf{y} \in A$, *il segmento* $[\mathbf{x}, \mathbf{y}]$ *è contenuto in* A.

***Esempio* 2.18** - In $\mathbb{R}$ gli insiemi convessi sono *tutti* e *soli* gli intervalli (limitati e non, aperti o chiusi o semiaperti etc. ...).

Evidentemente ogni insieme convesso di $\mathbb{R}^n$ è connesso per segmenti (e dunque anche connesso). È importante osservare che in $\mathbb{R}$ vale anche il viceversa; per cui, tenuto conto che gli insiemi convessi di $\mathbb{R}$ sono solo gli intervalli, possiamo affermare quanto segue.

▪ **Teorema 2.6** - *In* $\mathbb{R}$ *gli insiemi connessi sono tutti e solo gli intervalli.*

Esercizi

1. Consideriamo il sottoinsieme E di $\mathbf{R}^2$ definito da

$$E = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 > 1\} \cup \{(x, y) \in \mathbb{R}^2 : xy = 0\} .$$

Determinarne: punti interni, esterni, di frontiera, di accumulazione.

2. Considerare

a) $\mathbb{Z} \subset \mathbb{R}$

b) $\mathbb{Z} \subset \mathbf{R}^*$.

Verificare nei due casi se $\mathbb{Z}$ è chiuso e/o limitato.

3. Dimostrare che se $E \subset \mathbb{R}$ è limitato superiormente e $\sup E \notin E$, allora $\sup E$ è un punto di accumulazione di E.

4. Sia E compatto in $\mathbb{R}^n$. Se E è infinito allora E ha un punto di accumulazione x_0 ed inoltre $x_0 \in E$.
Vero o falso?

5. Siano $E = (0, 1)$ ed $E_n = (1/n, 1 - 1/n)$.

Dimostrare che $\bigcup_{n=3}^{\infty} E_n = E$ e quindi che $\{E_n\}_{n \geq 3}$ è una copertura di E, ma che

non ha sottocoperture finite.

6. Siano $A, B \subseteq \mathbb{R}^n$. Vero o falso:

a) A compatto e B chiuso $\Longrightarrow A \cap B$ compatto.

b) A aperto e B chiuso $\Longrightarrow A \cap B$ chiuso.

c) A e B aperti con $B \subset A \Longrightarrow \partial B \subset A$.

d) A aperto, B chiuso con $B \subset A \Longrightarrow \partial B \subset A$.

7. Siano $A, B \subseteq \mathbb{R}^n$. Vero o falso:

a) A e B connessi $\Longrightarrow A \cup B$ connesso.

b) A e B connessi $\Longrightarrow A \cap B$ connesso.

c) A e B convessi $\Longrightarrow A \cap B$ convesso.

d) A connesso per segmenti e B connesso $\Longrightarrow A \cap B$ connesso per segmenti.

8. Stabilire quale/i tra gli insiemi seguenti è connesso:

a) $\{x \in \mathbb{R}: |x| > 1\}$

b) $\{x \in \mathbb{R}^*: |x| > 1\}$

c) $\{x \in \dot{\mathbb{R}}: |x| > 1\}$.

L'operazione di limite 4

Dopo aver presentato gli spazi $\mathbb{R}^n$, il nostro scopo è quello di studiare le applicazioni tra di essi e, per cominciare, le applicazioni di $\mathbb{R}$ in $\mathbb{R}$, cioè le funzioni a valori reali di variabile reale.

La ragione per cominciare lo studio delle applicazioni da questo caso particolare è duplice: innanzitutto molti concetti sono più facilmente comprensibili, perché più intuitivi, essendo sorretti dalla rappresentazione grafica; inoltre, a causa dell'ordinamento di $\mathbb{R}$, sussistono molti risultati che sono peculiari di queste applicazioni, e non valgono nel caso più generale di funzioni di più variabili (cioè applicazioni da $\mathbb{R}^n$ a $\mathbb{R}$) o addirittura di applicazioni da $\mathbb{R}^n$ a $\mathbb{R}^m$.

Ma con le sole operazioni consentite in $\mathbb{R}$ dalla sua struttura algebrica, questo studio risulterebbe assai povero (dal punto di vista dell'Analisi). Utilizzeremo allora la struttura topologica di $\mathbb{R}$ e dei suoi ampliamenti $\mathbb{R}^*$ ed $\dot{\mathbb{R}}$, per introdurre un'operazione di tipo nuovo, del tutto diversa da quelle che finora conosciamo: l'*operazione* di **limite**. Essa costituirà lo strumento principale per affrontare lo studio delle funzioni. Cominceremo col definire questa operazione per le funzioni da $\mathbb{R}$ in $\mathbb{R}$, illustrando i suoi rapporti con l'ordinamento di $\mathbb{R}$ e le operazioni algebriche (calcolo dei limiti). Esamineremo poi, sia per la ricchezza dei risultati particolari che per l'importanza nei successivi sviluppi, il caso delle *successioni*. Estenderemo infine questa operazione alle applicazioni da $\mathbb{R}^n$ a $\mathbb{R}^m$ e in $\mathbb{C}$.

1. FUNZIONI REALI DI VARIABILE REALE

1.1 Positività e simmetrie

Iniziamo lo studio delle funzioni $x \mapsto f(x)$ dove sia la variabile indipendente x sia la variabile dipendente $y = f(x)$ appartengono a $\mathbb{R}$. Avremo sott'occhio il corredo di esempi illustrati nei diagrammi da 1 a 42 delle tavole fuori testo. Indicheremo generalmente con X il dominio di definizione di f, $\mathbb{R}$ sarà il codominio, $f(X) \subseteq \mathbb{R}$ l'immagine. Se f_1 e f_2 sono funzioni definite sullo stesso insieme X, e c_1 e c_2 numeri reali, risulterà definita su X ogni combinazione lineare $c_1 f_1 + c_2 f_2$ (cioè $x \mapsto c_1 f_1(x) + c_2 f_2(x)$); in particolare risulteranno definite su X le funzioni $f_1 + f_2$, $f_1 - f_2$; inoltre risulterà definita su X anche $f_1 \cdot f_2$ (cioè $x \mapsto f_1(x) \cdot f_2(x)$)

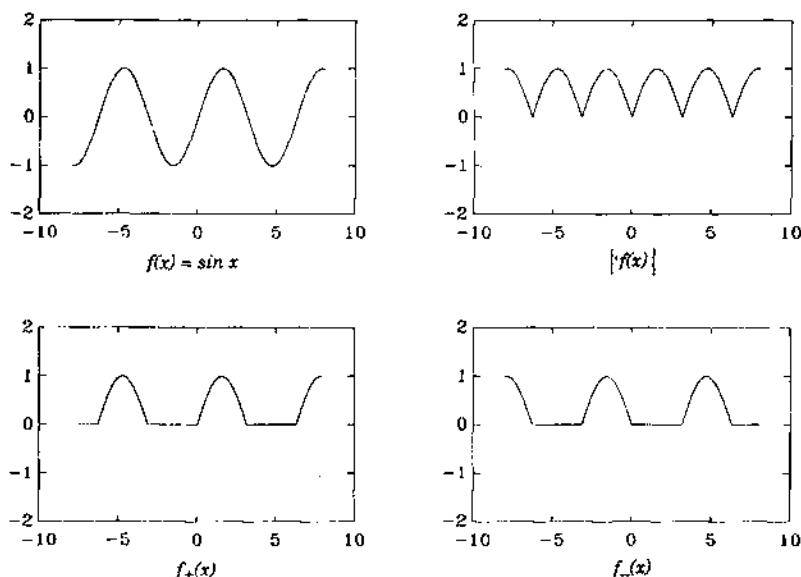



Fig. 4.1

mentre f_1/f_2 risulterà definita nei punti di X in cui sia $f_2(x) \neq 0$.

Una funzione $f : X \to \mathbb{R}$ si dirà *positiva* (*negativa*) se la sua immagine $f(X)$ è contenuta in $\mathbb{R}_+$ ($\mathbb{R}_-$). Una funzione positiva (negativa) ha il grafico contenuto nel semipiano superiore (inferiore).

Accanto a f risulterà definita anche la funzione *valore assoluto* di f, $|f|$, il cui grafico si ottiene da quello di f ribaltando al di sopra dell'asse delle ascisse le parti del grafico di f che ne stanno al di sotto (cfr. Fig. 4.1).

Risultano anche definite le funzioni *parte positiva* di f, f_+, e *parte negativa* di f, f_-, nel modo seguente

$$f_+(x) := \frac{1}{2}(|f(x)| + f(x)) = \max\{f(x), 0\}$$

$$f_-(x) := \frac{1}{2}(|f(x)| - f(x)) = \max\{-f(x), 0\} ,$$

cosicché risulta

$$f(x) = f_+(x) - f_-(x) \qquad |f(x)| = f_+(x) + f_-(x) .$$

Si noti che, sia la parte positiva f_+ che la parte negativa f_- sono funzioni non negative! (cfr. Fig. 4.1).

Una funzione (che per semplicità riteniamo definita su tutto l'asse reale) si dice *pari* quando è $f(-x) = f(x)$, *dispari* quando è $f(-x) = -f(x)$, per ogni $x \in \mathbb{R}$.

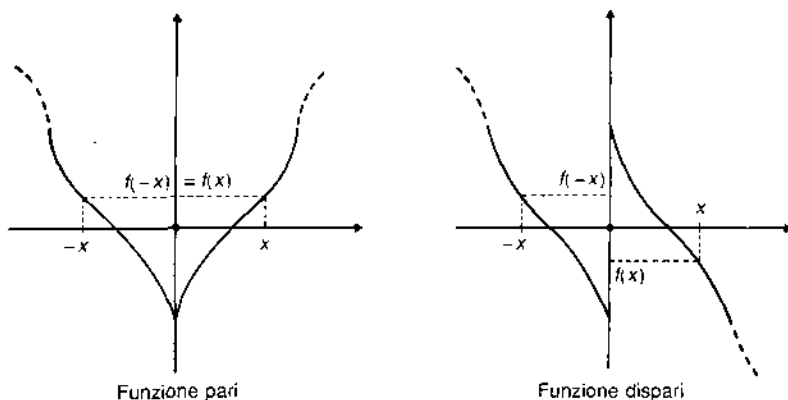



Fig. 4.2

Il grafico di una funzione pari è simmetrico ripetto all'asse delle y; quello di una funzione dispari è simmetrico rispetto all'origine degli assi (cfr. Fig. 4.2). Per esempio, ogni potenza pari $x^{2n}(n \in \mathbb{N})$ è una funzione pari, ogni potenza dispari x^{2n+1} è una funzione dispari. Ogni funzione $f(x)$ (definita su tutto $\mathbb{R}$) può essere decomposta nella somma di due funzioni, una pari e l'altra dispari, nel modo seguente:

$$f(x) = f_p(x) + f_d(x)$$

dove $f_p(x) := (f(x) + f(-x))/2$ è pari e $f_d(x) := (f(x) - f(-x))/2$ è dispari.

Il grafico di una funzione può presentare simmetrie meno elementari di quelle sopra segnalate. Una simmetria notevole è quella rispetto alla bisettrice del 1° e 3° quadrante. Si osservi infatti che se f è invertibile (cioè iniettiva), il grafico di f^{-1} si ottiene semplicemente riflettendo il grafico di f rispetto alla detta bisettrice (cfr. x^3 e $\sqrt[3]{x}$).

Allora, se il grafico di f è simmetrico rispetto a questa bisettrice, significa che la funzione è invertibile e coincide con la sua inversa (cfr. Fig. 4.5).

1.2 Funzioni limitate

Una funzione $f : X \to \mathbb{R}$ si dice *limitata* (*limitata inferiormente, superiormente*) se l'immagine $f(X)$ è un insieme limitato (risp. limitato inferiormente, superiormente) in $\mathbb{R}$.

In altre parole, f è limitata (superiormente, inferiormente), se esiste $M \in \mathbb{R}$, tale che

$$|f(x)| < M \ \text{(risp.} \ f(x) < M, \ f(x) > M) \ \forall x \in X \ .$$

Geometricamente, il grafico di una funzione *limitata* è contenuto in una striscia orizzontale del piano delimitata dalle rette $y = M$ e $y = -M$.

Si chiama *estremo superiore (inferiore)* di f l'estremo superiore (inferiore) dell'immagine $f(X)$ e si scrive semplicemente

$$\sup_X f \quad \text{o anche} \quad \sup f \quad \text{in luogo di} \quad \sup_{x\in X} f(x)$$

$$\inf_X f \quad \text{o anche} \quad \inf f \quad \text{in luogo di} \quad \inf_{x\in X} f(x)\,.$$

Se f è limitata, sia $\sup f$ che $\inf f$ stanno in $\mathbb{R}$; altrimenti sarà $\sup f = +\infty$ (se f non è limitata superiormente), $\inf f = -\infty$ (se f non è limitata inferiormente).

Ricordando le proprietà $i)$ e $ii)$ di 2.4 che definiscono l'estremo superiore in $\mathbb{R}$, possiamo caratterizzare l'estremo superiore di una funzione $f : X \to \mathbb{R}$ limitata superiormente nella maniera seguente:

$i)$ $\forall x \in X\ ,\ f(x) \leq \sup_X f$

$ii)$ $\forall \varepsilon > 0\ ,\ \exists x_\varepsilon \in X$ tale che: $f(x) > \sup_X f - \varepsilon$.

Analoga è la caratterizzazione di $\inf_X f$.

Si dirà *massimo* (*minimo*) di f il massimo (minimo) di $f(X)$, se esistono; li indicheremo con $\max f, \min f$. Si diranno invece *punti di massimo (minimo)* quei punti $x \in X$ tali per cui $f(x)$ è massimo (minimo). Tali punti si dicono anche *estremanti*, mentre si dicono *estremi* di f il massimo ed il minimo.

È evidente che, se f è limitata superiormente e ammette massimo, allora $\max f = \sup f$; mentre se f non è limitata superiormente, cioè $\sup f = +\infty$, allora essa non ammette massimo; ma ci sono anche funzioni limitate superiormente che non hanno massimo. Osservazioni analoghe valgono per le funzioni limitate inferiormente e per il minimo.

Esempi

1.1. Per la funzione $x \mapsto \sin x$, da $\mathbb{R}$ in $\mathbb{R}$, 1 è massimo, -1 è minimo; $\pi/2+2k\pi$ (con $k \in \mathbb{Z}$) sono punti di massimo, $-\pi/2 + 2k\pi$ sono punti di minimo.

1.2. La funzione $f(x) = (x) := x - [x]$ (mantissa di x) è limitata poiché $0 \leq f(x) < 1$ per ogni $x \in \mathbb{R}$; risulta $\inf f = \min f = 0$ e ogni intero relativo è punto di minimo; risulta anche $\sup f = 1$ (infatti, fissato $\varepsilon > 0$, risulta $f(x) > 1-\varepsilon$ se, per esempio, $1 - \varepsilon < x < 1$) ma f non ha massimo.

1.3. La funzione $f(x) = x \sin x$ non è limitata e si ha $\sup f = +\infty$ e $\inf f = -\infty$; infatti basta osservare che, per ogni $n \in \mathbb{N}$, $f((4n+1)\pi/2) = (4n+1)\pi/2$ e perciò, al crescere di n, l'immagine di $(4n+1)\pi/2$ supera ogni numero prefissato; analogamente si ragiona per l'estremo inferiore.

1.4. $x \mapsto \log_a x$ $(a > 0, a \neq 1)$ da $\mathbb{R}_+$ su $\mathbb{R}$ non è limitata, né inferiormente, né superiormente; $x \mapsto a^x$, da $\mathbb{R}$ in $\mathbb{R}_+$ è limitata inferiormente (da 0), ma non superiormente (basta rivedere la definizione e le proprietà di queste funzioni in 2.3.2, 2.3.3).

Si noti che, poiché consideriamo funzioni da $\mathbb{R}$ in $\mathbb{R}$, sia gli estremi di f che gli

estremanti sono elementi di $\mathbb{R}$, mentre $\sup f$ e $\inf f$ sono elementi di $\mathbb{R}^*$. Se la funzione è limitata, allora anche $\sup f$ e $\inf f$ appartengono a $\mathbb{R}$. Essi, come subito si riconosce dalla definizione, o sono punti isolati di $f(X)$ (in tal caso appartengono a $f(X)$ e perciò sono massimo e minimo di f) o sono punti di accumulazione di $f(X)$ (cfr. anche l'Osservazione 3.2.3).

I concetti di massimo e minimo sopra introdotti sono concetti *globali*, cioè riguardano la funzione nella sua totalità. Essi vengono chiamati anche *massimo globale o assoluto* (*minimo globale o assoluto*) per distinguerli dagli analoghi concetti *locali* che ora definiamo.

Definizione 1.1 - *Sia $f : X \to \mathbb{R}$ e $x_0 \in X$.*
Supponiamo che esista un intorno U di x_0 tale che si abbia

$$f(x) \geq f(x_0) \qquad se\ x \in U \cap X\ . \tag{1.1}$$

Allora $f(x_0)$ è un minimo locale o relativo, e x_0 è un punto di minimo locale relativo; se, invece della (1.1), *vale la*

$$f(x) > f(x_0) \quad per \quad x_0 \neq x \in U \cap X \tag{1.2}$$

allora si parla di minimo locale forte.

Analoga definizione per il *massimo locale (forte)* e *punto di massimo locale (forte)*.

I concetti ora introdotti, essendo relativi a un intorno di un punto del dominio di definizione di f, esprimono proprietà *locali* di f. Essi sono illustrati nel grafico seguente.

In riferimento alla figura 4.3: i punti a ed x_1 sono di minimo locale forte; i punti x tali che $x_2 < x < x_3$ sono sia di massimo sia di minimo locale debole; x_2 è di minimo locale debole, x_3 di massimo locale debole. I punti x_0 e b sono di massimo locale forte; x_0 è inoltre di massimo *globale*. La funzione ha dunque massimo $f(x_0)$, ma non ha minimo (globale). È tuttavia limitata inferiormente ed il suo estremo inferiore è l.

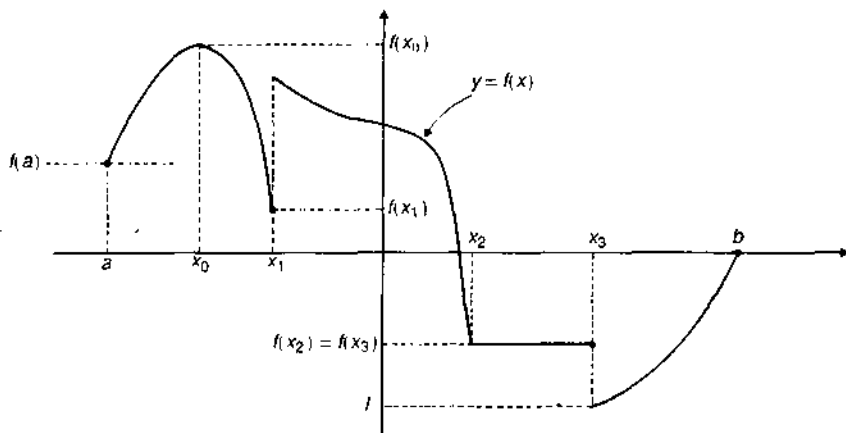



Fig. 4.3

***Esempio* 1.5** - La funzione $x \mapsto x \sin x$ (grafico n. 38), essendo non limitata né inferiormente né superiormente, non ha massimo né minimo; ma possiede infiniti minimi relativi (il maggiore dei quali è 0) e infiniti massimi relativi.

1.3 Funzioni monotone

Una funzione $f : X \to \mathbb{R}$ si dice *crescente* in X se, per ogni coppia di punti x_1, x_2 di X, con $x_1 > x_2$, risulta $f(x_1) \geq f(x_2)$; ciò si può anche esprimere dicendo : f è crescente in X se

$$\frac{f(x_1) - f(x_2)}{x_1 - x_2} \geq 0 \qquad \forall x_1, x_2 \in X,\ x_1 \neq x_2 \ . \tag{1.3}$$

Analoga definizione (col segno $\leq$ nella (1.3)) si darà per una funzione *decrescente*. Se nella (1.3) possiamo mettere il segno $>$ invece di $\geq$ diremo che f è *strettamente crescente* in X (*strettamente decrescente* se possiamo mettere il segno $<$) *.

Tutti questi tipi di funzioni (crescenti o decrescenti, strettamente o non) prendono l'appellativo comune di funzioni *monotone*. In particolare, le funzioni costanti rientrano in questa classe.

Esempi (potenze, esponenziali, logaritmi)

1.6. Le funzioni $x \mapsto x^\alpha$ $(\alpha > 0)$ sono tutte strettamente crescenti in $[0, +\infty)$; ciò discende dalla proprietà E_5 di 2.3.2. Esse sono biiettive da $[0, +\infty)$ in $[0, +\infty)$. Le loro inverse: $x \mapsto x^{1/\alpha}$ sono monotone dello stesso tipo.

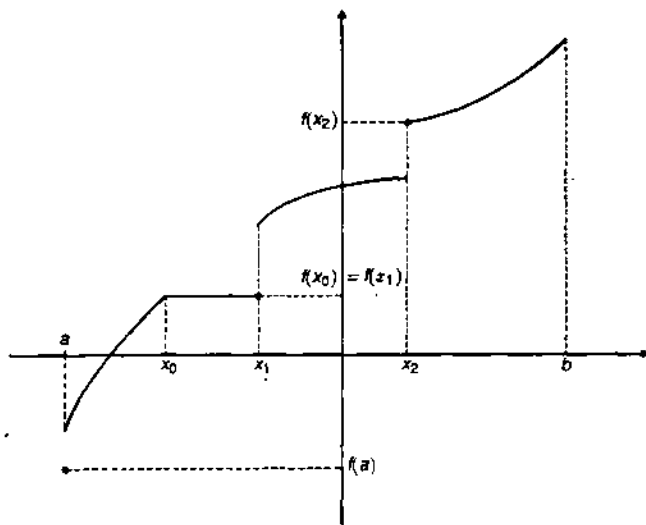



Fig. 4.4 Funzione monotona (non strettamente) definita nell'intervallo $[a, b)$.

* Invece di *crescente* (risp. *decrescente*) si può anche dire *non decrescente* (risp. *non crescente*).

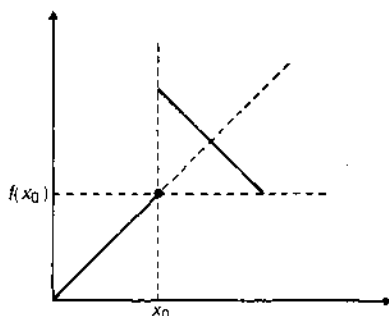



Fig. 4.5

1.7. Le funzioni: $x \mapsto a^x$ $(a > 0)$, definite su $\mathbb{R}$, sono strettamente crescenti per $a > 1$, strettamente decrescenti per $a < 1$; ciò segue dalla proprietà E_4 di 2.3.2. Esse sono biiettive tra $\mathbb{R}$ ed $\mathbb{R}_+$.

1.8. Le funzioni: $x \mapsto \log_a x$ $(a > 0, a \neq 1)$, definite in $\mathbb{R}_+$ sono strettamente crescenti se $a > 1$, strettamente decrescenti se $a < 1$; esse sono infatti le funzioni inverse di quelle dell'esempio 1.7.

Le proprietà sopra illustrate sono di tipo *globale*. Sarà utile considerare anche funzioni monotone solo localmente.

Sia dunque $f : X \to \mathbb{R}$ e $x_0 \in X$ un punto non isolato di X; diremo che f è *crescente nell'intorno di* x_0 se esiste un intorno U di x_0 per cui vale la (1.3) $\forall x_1, x_2 \in X \cap U, x_1 \neq x_2$. Analoghe definizioni sussistono negli altri casi.

***Esempio* 1.9** - La funzione 5: $x \mapsto [x]$ è crescente; la 6: $x \mapsto (x)$ non è monotona; però essa è strettamente crescente nell'intorno di ogni punto $x \neq 0, \pm 1, \pm 2, \ldots$

Segue subito dalla definizione che se $f : X \to \mathbb{R}$ è strettamente monotona in X, allora è iniettiva e quindi invertibile e la funzione inversa f^{-1} è monotona dello stesso tipo. La condizione di monotonia non è però necessaria per l'invertibilità, come dimostra la funzione il cui grafico è riportato in Fig. 4.5 (funzione non monotona, invertibile, coincidente con la sua inversa).

Le funzioni monotone, come vedremo, presentano un comportamento particolare rispetto all'operazione di limite che ora passiamo a definire.

Esercizi

1. Disegnare i grafici di $|f|, f_+, f_-$ per le seguenti funzioni

$$f(x) = \cos x, \;\; f(x) = \log_2 x, \;\; f(x) = x^3 \; .$$

2. Sia data $f : [-3, 2] \to \mathbb{R}$ il cui grafico è indicato in figura:

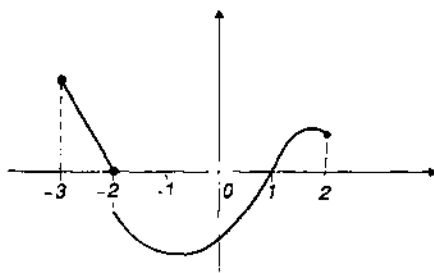



Fig. 4.6

Disegnare il grafico di: $-f, |f|, f_+, f_-, f(|x|)$.

3. Disegnare i grafici delle seguenti funzioni:

a) $\frac{1}{2}(|x| - x)$

b) $\max\{x^2, x^3\}$

c) $[x^2]$

d) $|x^2 - 1| + 2x$

e) $\sup_{t \in [x, +\infty)} \{-t^3 - t\}$

4. Trovare i domini di definizione naturali delle seguenti funzioni:

a) $\log_a |x| \qquad (a > 0, a \neq 1)$

b) $\log_a \log_b x \qquad (a, b > 0, \neq 1)$

c) $\sqrt[3]{1 - \sqrt{x^2 - 4}}$

5. Vero o falso:

a) una funzione pari non è invertibile;

b) una funzione dispari è invertibile.

6. Sia $f : X \to \mathbb{R}$. Vero o falso:

a) f limitata $\Rightarrow$ f ha massimo e minimo (globali).

b) f limitata $\Leftrightarrow$ $\sup f \in \mathbb{R}$ e $\inf f \in \mathbb{R}$.

c) $\sup f$ e $\inf f$ sono punti di accumulazione (in $\mathbb{R}^*$) di $f(X)$.

7. Trovare gli estremi superiore e inferiore delle seguenti funzioni, indicando se trattasi di massimi e minimi.

a) $(-\pi/2, \pi/2) \ni x \mapsto \operatorname{tg} x$

b) $[-\pi/4, \pi/4) \ni x \mapsto \operatorname{tg} x$

c) $[-1, +1] \ni x \mapsto \sqrt{1 - x^2}$

d) $x \mapsto \sin \frac{1}{x}$

8. Vero o falso:

a) f e g monotone $\Rightarrow f \cdot g$ e $f + g$ monotone.

b) f e g crescenti (decrescenti) $\Rightarrow f \cdot g$ e $f + g$ crescenti (decrescenti).

c) f crescente e g crescente $(g \neq 0) \Rightarrow f/g$ crescente.

d) siano f e g componibili:

f crescente (decrescente) e g crescente $\Rightarrow f \circ g$ crescente

f crescente (decrescente) e g decrescente $\Rightarrow f \circ g$ decrescente.

9. Verificare che la funzione definita da:

$$f(x) = \begin{cases} x & \text{se } x \in \mathbb{Q} \\ -x & \text{se } x \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

è invertibile ma non esiste alcun intervallo in cui è monotona.

2. LIMITI DI FUNZIONI DA $\mathbb{R}$ IN $\mathbb{R}$

2.1 Definizione di limite

Siano $X \subseteq \mathbb{R}$ e $f : X \to \mathbb{R}$. Scopo della nozione di limite è descrivere rigorosamente il comportamento di una funzione "vicino" ad un punto di accumulazione del suo dominio X.

Pensando a funzioni illimitate e/o definite su domini illimitati in $\mathbb{R}$, si comprende che l' "ambiente naturale" per tale definizione non è $\mathbb{R}$ bensì uno dei suoi ampliamenti $\mathbb{R}^*$ o $\dot{\mathbb{R}}$. Per funzioni reali di variabile reale risulta preferibile $\mathbb{R}^*$, che permette di sfruttare l'ordinamento. Daremo tuttavia anche la definizione in $\dot{\mathbb{R}}$. Dunque opereremo in $\mathbb{R}^*$ con la topologia che abbiamo definito in 3.2.3.

Sia ora $x_0 \in \mathbb{R}^*$ un punto di accumulazione per X (perciò ammettiamo che x_0 possa essere un numero reale, oppure $+\infty$ o $-\infty$). Possiamo allora "far tendere x a x_0" (in simboli: $x \to x_0$); con questa espressione si intende dire che è possibile prendere intorni sempre più piccoli di x_0 e sempre, in ognuno di essi, cadranno infiniti punti di X.

Ora, la questione principale è: cosa succede delle immagini $f(x)$ quando $x \to x_0$? Per illustrare compiutamente le varie situazioni che possono presentarsi, diamo anzitutto la seguente definizione.

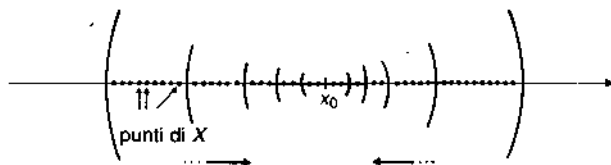



Fig. 4.7 Intorni che si stringono in x_0.

Definizione 2.1 - *Sia $\mathcal{P}$ una proprietà relativa alla funzione $f : X \to \mathbb{R}$; si dice che, per $x \to x_0$, f possiede (o acquista) definitivamente la proprietà $\mathcal{P}$ se esiste un intorno U di x_0 tale che nei punti $x \in X \cap U$ e distinti da x_0, la funzione f possiede la proprietà $\mathcal{P}$.* (Sottolineiamo che la proprietà $\mathcal{P}$ non è richiesta nel punto x_0, il quale potrebbe anche non appartenere a X).

Esempi

2.1. $1 - x^2$ è definitivamente positiva per $x \to 0$; infatti è positiva nell'intorno $(-1, 1)$.

2.2. $1/x$ è definitivamente < 1 per $x \to +\infty$; infatti è < 1 nell'intorno $(1, +\infty]$.

2.3. $|x|$ è definitivamente maggiore di x^2 per $x \to 0$; infatti risulta $|x| - x^2 > 0$ nell'intervallo $(-1, 1)$ privato dell'origine.

Definizione 2.2 - *Sia $f : X \to \mathbb{R}$ e $x_0 \in \mathbb{R}^*$ un punto di accumulazione per X; sia inoltre l un elemento di $\mathbb{R}^*$. Diremo che : "il* **limite** *di $f(x)$, per x che tende a x_0, è l ", o anche: "$f(x)$ tende a l per x tendente a x_0", e scriveremo*

$$\lim_{x \to x_0} f(x) = l \quad \textit{oppure} \quad f(x) \to l \quad \textit{per} \quad x \to x_0$$

se, per ogni intorno V di l, è possibile trovare un intorno U di x_0 per cui

$$f(x) \in V \quad \textit{se} \quad x_0 \neq x \in U \cap X \, .$$

In altre parole: ... *se, per $x \to x_0$, $f(x)$ appartiene definitivamente ad un qualsiasi intorno V di l.*

Osservazione 2.1 - Notiamo ancora che x_0 può non appartenere a X e perciò $f(x_0)$ può non essere definito; anche se è definito, $f(x_0)$ può non avere nulla a che vedere con il limite l.

Osservazione 2.2 - Un limite può dunque essere un numero, oppure $+\infty$ o $-\infty$, oppure può anche non esistere. Se esiste, esso può appartenere o non appartenere all'immagine di f; se non vi appartiene, è comunque un punto di accumulazione per questo insieme, come segue dalla definizione di limite.

A seconda che x_0 e l siano numeri oppure $+\infty$ o $-\infty$ si hanno vari casi particolari. Ripetiamo esplicitamente la definizione in qualche caso, invitando l'allievo ad esplicitare *tutti* i casi particolari.

1. Siano $x_0, l \in \mathbb{R}$. Allora un intorno V di l sarà del tipo $(l - \varepsilon, l + \varepsilon)$ con $\varepsilon > 0$ e un intorno U di x_0 sarà $(x_0 - \delta, x_0 + \delta)$ con $\delta > 0$. La definizione prende la forma seguente: $\lim_{x \to x_0} f(x) = l$ se, per ogni numero positivo ε, si può in corrispondenza trovare un numero positivo δ (dipendente da $\varepsilon : \delta = \delta(\varepsilon)$) in modo tale che:

$$|f(x) - l| < \varepsilon \quad \text{se} \quad 0 < |x - x_0| < \delta \quad \text{e} \quad x \in X \, .$$

2. Sia $x_0 = -\infty$, $l \in \mathbb{R}$. Allora V sarà ancora del tipo $(l-\varepsilon, l+\varepsilon)$ con $\varepsilon > 0$, mentre un intorno U di $-\infty$ sarà del tipo $[-\infty, -k)$ con $k > 0$ (non è restrittivo considerare solo intorni con estremo destro negativo). Allora: $\lim_{x\to-\infty} f(x) = l$ se, per ogni $\varepsilon > 0$, si può in corrispondenza trovare un numero $k > 0$ $(k = k(\varepsilon))$ tale che:

$$|f(x) - l| < \varepsilon \quad \text{se} \quad x < -k \ \text{ e } \ x \in X \ .$$

3. Sia $x_0 \in \mathbb{R}$, $l = +\infty$. Allora $V = (k, +\infty]$ con $k > 0$ e $U = (x_0-\delta, x_0+\delta)$. Diremo: $\lim_{x\to x_0} f(x) = +\infty$ se, per ogni $k > 0$ si può in corrispondenza trovare un $\delta > 0$ $(\delta = \delta(k))$ tale che

$$f(x) > k \quad \text{se} \quad 0 < |x - x_0| < \delta \ \text{ e } \ x \in X \ .$$

4. Sia $x_0 = +\infty$, $l = -\infty$. Allora $V = [-\infty, -h)$ e $U = (k, +\infty]$ con h e $k > 0$. Diremo: $\lim_{x\to+\infty} f(x) = -\infty$ se, $\forall h > 0$, $\exists k > 0$ $(k = k(h))$ tale che:

$$f(x) < -h \quad \text{se} \quad x > k \ \text{ e } \ x \in X \ .$$

Le situazioni 1., 2., 3., 4. sono sommariamente illustrate nei seguenti diagrammi.

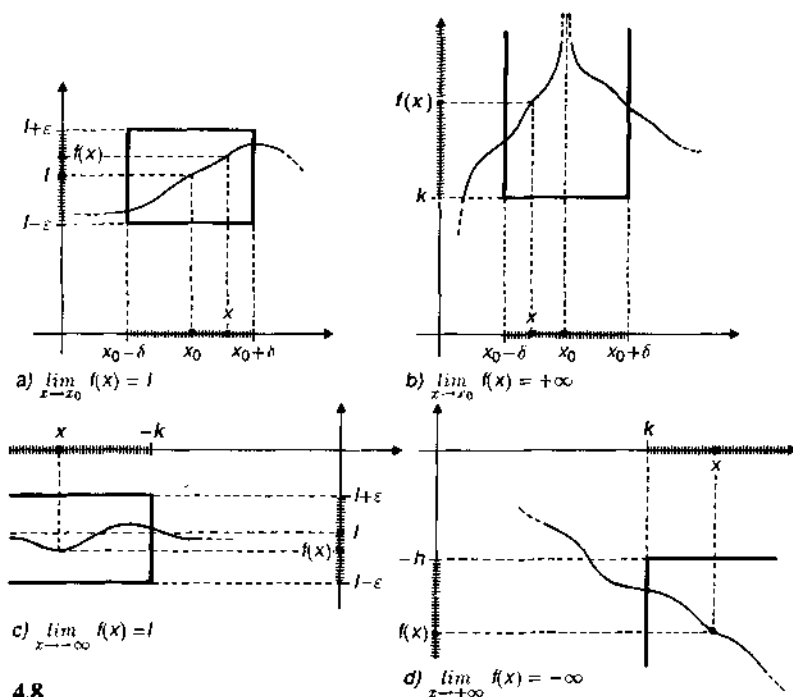



Fig. 4.8

Esempi

2.4. $\lim_{x \to 0} \sin x = 0$. Infatti, dalla trigonometria sappiamo che $|\sin x| < |x|$ (per $x \neq 0$) e quindi, per avere $|\sin x| < \varepsilon$ basta prendere $|x| < \varepsilon$ (cioè scegliere $\delta = \varepsilon$).

2.5. $\lim_{x \to 0} \cos x = 1$. Infatti, ricordando le formule di bisezione della trigonometria, abbiamo: $1 - \cos x = 2(\sin x/2)^2 < 2\dfrac{x^2}{4}$ e perciò sarà $1 - \cos x < \varepsilon$ se $|x| < \sqrt{2\varepsilon}$; sceglieremo allora $\delta = \sqrt{2\varepsilon}$.

2.6. $\lim_{x \to 0} \dfrac{1}{x^2} = +\infty$. Per avere $\dfrac{1}{x^2} > k$ basta prendere $|x| < 1/\sqrt{k}$, cioè scegliere $\delta = 1/\sqrt{k}$.

2.7. $\lim_{x \to +\infty} \sin x$ non esiste. Provarlo per esercizio (cfr. anche la seguente Osservazione 2.5).

L'operazione di limite non è sempre possibile. Ma, quando è possibile, è ben definita, cioè il limite, se esiste, è *unico*, come prova il seguente teorema.

■ **Teorema 2.1** - (Unicità del limite) *Sia*

$$\lim_{x \to x_0} f(x) = l_1 \qquad e \qquad \lim_{x \to x_0} f(x) = l_2 \, ;$$

allora $l_1 = l_2$.

Dimostrazione - Per assurdo, sia $l_1 \neq l_2$. Presi allora due intorni V_1 di l_1 e V_2 di l_2 disgiunti, dovrebbero esistere due intorni U_1 e U_2 di x_0 per cui

$$f(x) \in V_1 \quad \text{se} \quad x \in U_1 \cap X$$

e $\qquad (x \neq x_0)$

$$f(x) \in V_2 \quad \text{se} \quad x \in U_2 \cap X .$$

Dunque, prendendo $x \in U_1 \cap U_2 \cap X$ dovrebbe essere contemporaneamente $f(x) \in V_1$ e $f(x) \in V_2$; assurdo. □

Si osservi che la semplice dimostrazione discende dalla proprietà $\mathcal{U}_4$ degli intorni (proprietà di separazione; cfr. 3.1.4).

2.2 Limite destro, sinistro, per eccesso, per difetto; in $\dot{\mathbb{R}}$

La funzione n. 2: $x \mapsto \operatorname{sgn} x$ chiaramente non ammette limite per $x \to 0$. Ma questa è un'informazione piuttosto scarna circa il comportamento della funzione in un intorno dell'origine. Per avere un'informazione più precisa, introduciamo il concetto di **limite destro** e **limite sinistro**. Non è una definizione di limite diversa da quella sopra introdotta, ma sono diversi gli intorni (cioè la topologia) utilizzati.

Dato $x \in \mathbb{R}$, definiamo **intorno destro** di x qualsiasi intervallo $[x, x+r)$ con $r > 0$ e **intorno sinistro** qualsiasi intervallo $(x-r, x]$.

La famiglia degli intorni destri, così come quella degli intorni sinistri, verifica le proprietà $(\mathcal{U})$ di 3.1.4; cosicché è una topologia per $\mathbb{R}$. Gli intorni di $+\infty$ sono già intorni sinistri, mentre quelli di $-\infty$ sono già intorni destri. Se sull'insieme di definizione della funzione si prende questa topologia (degli intorni destri (rispett. sinistri)), mentre sul codominio si mantiene la solita, si ha il *limite destro* (rispett. *sinistro*) che si indicherà così:

$$\lim_{x \to x_{0+}} \qquad \text{rispettivamente} \qquad \lim_{x \to x_{0-}}$$

Cioè, $\lim\limits_{x \to x_{0+}} f(x) = l$ significa che, per ogni intorno V di l è possibile trovare un intorno destro U_+ di x_0 per cui

$$f(x) \in V \qquad \text{se} \qquad x_0 \neq x \in U_+ \cap X \ .$$

Analoga definizione per il *limite sinistro*. Si capisce che il $\lim\limits_{x \to +\infty}$ è un limite sinistro, il $\lim\limits_{x \to -\infty}$ è un limite destro.

***Esempio* 2.8** - $\lim\limits_{x \to 0_+} \operatorname{sgn} x = 1$, $\lim\limits_{x \to 0_-} \operatorname{sgn} x = -1$. Si noti che i due limiti, destro e sinistro, esistono, sono diversi fra loro (e diversi dal valore che la funzione assume in 0).

■ **Teorema 2.2** - *Condizione necessaria e sufficiente perché esista il* $\lim\limits_{x \to x_0} f(x)$ *è che esistano il limite destro* $(x \to x_{0+})$ *e sinistro* $(x \to x_{0-})$ *e siano uguali.*

La semplice dimostrazione viene lasciata per esercizio.

Se invece si prende sul codominio la topologia degli intorni destri (rispett. sinistri) e sul dominio di definizione la topologia solita si ha il **limite per eccesso** (rispett. **per difetto**) che si indicherà così:

$$\lim_{x \to x_0} f(x) = l_+ \quad \text{rispett.} \quad \lim_{x \to x_0} f(x) = l_- \ .$$

Cioè, $\lim\limits_{x \to x_0} f(x) = l_+$ significa che, per ogni intorno destro V_+ di l è possibile trovare un intorno U di x_0, per cui

$$f(x) \in V_+ \quad \text{se} \quad x_0 \neq x \in U \cap X \ .$$

Analoga definizione per il limite per difetto. È anche chiaro il significato di una scrittura del tipo

$$\lim_{x \to x_{0+}} f(x) = l_-$$

e di altre simili.

Esempi

2.9. $\lim_{x\to+\infty} x/(x+1) = 1_-$. Infatti, preso un intorno sinistro di 1, $(1-\varepsilon, 1]$ (con $\varepsilon > 0$) risulterà $1-\varepsilon < x/(x+1) \le 1$ per $x > (1-\varepsilon)/\varepsilon$, cioè per x appartenente all'intorno di $+\infty$ $(k, +\infty]$ con $k = (1-\varepsilon)/\varepsilon$.

2.10. $\lim_{x\to 1_+} 2^x = 2_+$. Infatti, fissato $\varepsilon > 0$, risulterà $2 \le 2^x < 2+\varepsilon$ per $1 \le x < \log_2(2+\varepsilon) = \log_2[2(1+\varepsilon/2)] = 1 + \log_2(1+\varepsilon/2)$.

Se anziché $\mathbb{R}^*$ si sceglie $\dot{\mathbb{R}}$ come ampliamento di $\mathbb{R}$ la definizione di limite 2.2 rimane inalterata nella sua struttura. Gli unici cambiamenti sono che x_0 ed l appartengono a $\dot{\mathbb{R}}$ e gli intorni vanno intesi nel senso della topologia in questo spazio.

Sarà dunque chiaro il significato di formule come le seguenti:

$$\lim_{x\to\infty} f(x) = l \in \mathbb{R} \ , \ \lim_{x\to x_0} f(x) = \infty \ (x_0 \in \mathbb{R}) \text{ oppure } \lim_{x\to\infty} f(x) = \infty \ .$$

Il seguente semplice esempio illustra la differenza tra $\mathbb{R}^*$ e $\dot{\mathbb{R}}$.

Si voglia calcolare $\lim_{x\to 0} 1/x$; poiché $\lim_{x\to 0_+} 1/x = +\infty$ e $\lim_{x\to 0_-} 1/x = -\infty$ il detto limite non esiste in $\mathbb{R}^*$. Ma basterà utilizzare, sul codominio, la topologia $\dot{\mathbb{R}}$ (cioè cercare il limite in $\dot{\mathbb{R}}$ invece che in $\mathbb{R}^*$) per trovarlo.

Risulta: $\lim_{x\to 0} 1/x = \infty$; infatti fissato un intorno di ∞ $\{y : |y| > k\}$ (con $k > 0$) risulterà $1/x$ appartenente a questo intorno se $x \in (-\delta, +\delta)$ con $\delta = 1/k$ $(x \ne 0)$.

Esercizi

1. Sia $f : X \to \mathbb{R}$; $x_0, l \in \mathbb{R}$.
Quale dei tre enunciati seguenti è la negazione del predicato: $\lim_{x\to x_0} f(x) = l$?

a) $\forall\varepsilon > 0$, $\exists\delta = \delta(\varepsilon)$ tale che: $\forall x \in X$, $0 < |x - x_0| < \delta \Rightarrow |f(x) - l| > \varepsilon$.

b) $\exists\varepsilon_0 > 0$ tale che: $\forall\delta > 0$, $\exists\overline{x} \in X$ tale che: $0 < |\overline{x} - x_0| < \delta$ e $|f(\overline{x}) - l| > \varepsilon_0$.

c) $\exists\varepsilon_0 > 0$ e $\delta_0 > 0$ tale che: $x \in X$ e $0 < |x - x_0| < \delta_0 \Rightarrow |f(x) - l| > \varepsilon_0$.

2. Sia $f : \mathbb{R} \to \mathbb{R}$. Negare che: $\lim_{x\to-\infty} f(x) = +\infty$ equivale ad affermare che:

a) $\exists M_0$ tale che: $\forall N > 0$, $\exists\overline{x} \in X$ tale che: $\overline{x} < -N$ e $f(x) < M_0$

b) $\forall M, \exists N = N(M)$ tale che: $\forall x \in X$, $x < -N \Rightarrow f(x) < M$.

c) Esistono $M_0 > 0$ ed $N_0 > 0$ tali che: $\forall x \in X$, $x < -N_0 \Rightarrow f(x) < M_0$.

3. Disegnare diagrammi come quelli della figura 4.8 che illustrino le seguenti situazioni:

a) $\lim_{x\to 3_+} f(x) = -\infty$

b) $\lim_{x\to+\infty} f(x) = 1_-$

c) $\lim_{x\to-\infty} f(x) = 1$ (né per difetto né per eccesso).

4. Verificare con la definizione che

a) $\lim_{x\to 0_\pm} 1/x^3 = \pm\infty$

b) $\lim_{x\to-\infty} 1/(x^2+1) = 0$

c) $\lim_{x\to 1_\pm} 1/(1-x) = \mp\infty$.

5. Dimostrare che se $l \in \mathbb{R}$

$$\lim_{x\to x_0} f(x) = l \quad \text{se e solo se} \quad \lim_{x\to x_0} |f(x) - l| = 0 .$$

In particolare $f(x) \to 0$ se e solo se $|f(x)| \to 0$.

6. La struttura della definizione di limite prevede di scegliere *arbitrariamente* un intorno di l e di *conseguenza* determinare un intorno di x_0 tale che...

Questo procedimento sembra andare contro l'intuizione; infatti parrebbe naturale scegliere *prima* l'intorno di x_0 e *poi* quello di l, in quanto si "immagina" *prima* di far tendere x ad x_0 e *poi* di osservare il comportamento di $f(x)$. Proviamo allora a dare la seguente definizione: $\lim_{x\to x_0} f(x) = l$ se, fissato arbitrariamente un intorno U di x_0, si può trovare di conseguenza un intorno V di l tale che: $f(x) \in V$ se $x \in U$.

Vi sembra significativa tale definizione?

2.3 Calcolo dei limiti

Lo studente avrà ben compreso a questo punto come l'operazione ora introdotta sia strettamente legata alla topologia utilizzata sul dominio e codominio delle funzioni. Vediamo ora come si comporta questa operazione rispetto all'ordinamento in $\mathbb{R}^*$ e alla struttura algebrica di $\mathbb{R}$.

Siano $f : X \to \mathbb{R}$; $x_0, l \in \mathbb{R}^*$ con x_0 punto di accumulazione per X.

■ **Teorema 2.3** - (Permanenza del segno). *Se* $\lim_{x\to x_0} f(x) = l > 0$ (< 0), *allora è definitivamente* $f(x) > 0$ (< 0) *per* $x \to x_0$.

Dimostrazione - Sia l finito e positivo. Preso un intorno V di l, diciamo $(l-\varepsilon, l+\varepsilon)$ con $0 < \varepsilon < l$, esiste un intorno U di x_0 tale che, se $x_0 \neq x \in U \cap X$, allora $f(x) \in V$, cioè $l + \varepsilon > f(x) > l - \varepsilon > 0$.

Analoga è la dimostrazione nei casi $l = +\infty$, l finito e negativo ed $l = -\infty$. □

Osservazione 2.3 - L'implicazione del teorema non è invertibile. Tutto ciò che si può affermare, se $f(x) > 0$ in un intorno di x_0 (x_0 escluso), è che il limite, *se esiste*, è ≥ 0.

Osservazione 2.4 - Con dimostrazione pressocché identica si vede che, *se* $f(x)$ *ammette limite finito per* $x \to x_0$, *allora* f *è localmente limitata, cioè esiste un*

intorno U di x_0 per cui $f(X \cap U)$ è limitato in $\mathbb{R}$. Anche questa affermazione non è invertibile, poiché una funzione localmente limitata in x_0 può non ammettere limite per $x \to x_0$. Per esempio, $\sin x$ non ha limite per $x \to +\infty$ pur essendo limitata.

■ **Teorema 2.4** - (del confronto). *Siano f, g, h funzioni da X in $\mathbb{R}$ e $x_0 \in \mathbb{R}^*$ un punto di accumulazione per X.*

Se $\lim\limits_{x \to x_0} f(x) = \lim\limits_{x \to x_0} h(x) = l \in \mathbb{R}^*$ e se esiste un intorno U di x_0 tale che risulti

$$f(x) \leq g(x) \leq h(x)$$

per ogni $x_0 \neq x \in X \cap U$, allora è anche $\lim\limits_{x \to x_0} g(x) = l$.

Dimostrazione - Sia $l \in \mathbb{R}$; preso un intorno V di l, $(l - \varepsilon, l + \varepsilon)$ esistono intorni U_1 e U_2 di x_0 per cui, quando $x \in U_1 - \{x_0\}$, $f(x) \in V$ e quando $x \in U_2 - \{x_0\}$, $h(x) \in V$; allora, preso un intorno U di x_0 contenuto in $U_1 \cap U_2$, succede che: $l - \varepsilon \leq f(x) \leq g(x) \leq$ $\leq h(x) \leq l + \varepsilon$, cioè $g(x) \in V$ per $x \in U - \{x_0\}$, da cui la tesi.

Analoga è la dimostrazione se $l = +\infty$ (in tal caso è sufficiente l'ipotesi $g(x) \geq f(x)$) o $l = -\infty$ (in tal caso è sufficiente che $g(x) \leq h(x)$). □

Ecco un notevole esempio di applicazione del teorema 2.4.

***Esempio* 2.11** - Vogliamo provare che:

$$\lim_{x \to 0} \frac{\sin x}{x} = 1 \ . \tag{2.1}$$

Con riferimento alla Fig. 4.9, sia x $(0 < x < \pi/2)$ la lunghezza dell'arco PA (misurata in radianti) della circonferenza con centro in O e raggio 1. Allora la lunghezza del segmento PH è $\sin x$ e quella del segmento QA è $\operatorname{tg} x$; l'area del triangolo OAP è $(\sin x)/2$, quella del settore circolare $\widehat{OAP}$ è $x/2$ e quella del triangolo OAQ è $(\operatorname{tg} x)/2$. Si ha dunque

$$\sin x < x < \operatorname{tg} x$$

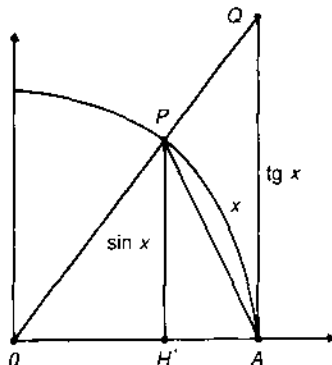



Fig. 4.9

da cui, dividendo per $\sin x$ e prendendo i reciproci,

$$\cos x < \frac{\sin x}{x} < 1 \,.$$

Queste disuguaglianze non cambiano scrivendo $-x$ al posto di x; perciò restano valide anche per $-\pi/2 < x < 0$. Abbiamo già mostrato (cfr. esempio 2.5) che $\lim_{x \to 0} \cos x = 1$; perciò, dal teorema 2.4, segue la (2.1).

Veniamo ora al cosiddetto *calcolo dei limiti*, cioè alle proprietà dell'operazione di limite rispetto alle operazioni algebriche. Cominciamo con un caso semplice.

■ **Teorema 2.5** - *Supponiamo che esistano in* $\mathbb{R}$ *(cioè esistano finiti) i limiti seguenti:*

$$\lim_{x \to x_0} f(x) = l_1 \ , \ \lim_{x \to x_0} g(x) = l_2$$

dove $x_0 \in \mathbb{R}^*$. *Allora risulta, per* $x \to x_0$,

i)	$cf(x)$	$\to$	cl_1	$\forall c \in \mathbb{R}$
ii)	$f(x) \pm g(x)$	$\to$	$l_1 \pm l_2$	
iii)	$f(x)g(x)$	$\to$	$l_1 l_2$	
iv)	$1/g(x)$	$\to$	$1/l_2$	*(se* $l_2 \neq 0$*)*
v)	$f(x)/g(x)$	$\to$	l_1/l_2	*(se* $l_2 \neq 0$*)* .

Dimostrazione - Lasciamo per esercizio le *i)* e *ii)*. Per la *iii)* si osservi che:

$$f(x)g(x) - l_1 l_2 = g(x)(f(x) - l_1) + l_1(g(x) - l_2)$$

e perciò, fissato $\varepsilon > 0$, è definitivamente $|f(x) - l_1| < \varepsilon$, $|g(x) - l_2| < \varepsilon$, e inoltre $|g(x)| < k$ (per l'Osservazione 2.4), per cui sarà definitivamente

$$|f(x)g(x) - l_1 l_2| < k\varepsilon + l_1\varepsilon = (k + l_1)\varepsilon$$

e, per l'arbitrarietà di ε, segue la tesi.

Per provare la *iv)* osserviamo anzitutto che, essendo $l_2 > 0$ (per esempio) sarà definitivamente $g(x) > 0$, oltre che limitata, diciamo $0 < h \leq g(x) < k$ e perciò $1/k < 1/g(x) \leq$ $\leq 1/h$. Quindi abbiamo, fissato ε, definitivamente

$$\left| \frac{1}{g(x)} - \frac{1}{l_2} \right| = \frac{|l_2 - g(x)|}{|l_2 g(x)|} \leq \frac{\varepsilon}{|l_2| h}$$

da cui l'asserto.

La *v)* segue con considerazioni analoghe osservando che si può scrivere

$$\frac{f(x)}{g(x)} - \frac{l_1}{l_2} = \frac{l_2 f(x) - l_1 g(x)}{l_2 g(x)} = \frac{l_2(f(x) - l_1) - l_1(g(x) - l_2)}{l_2 g(x)}$$

Lasciamo la conclusione per esercizio. □

***Esempio* 2.12** - Vogliamo provare che

$$\lim_{x\to 0}\frac{1-\cos 2x}{x^2}=2\ . \tag{2.2}$$

Ricordando le formule di bisezione della trigonometria (cfr. esempio 2.5) abbiamo che:

$$\frac{1-\cos 2x}{x^2}=\frac{2(\sin x)^2}{x^2}=2\left(\frac{\sin x}{x}\right)^2 .$$

Applicando allora la *iii*) del teorema 2.5 (con $f(x)=g(x)=\sin x/x$) e ricordando il limite (1.1) si ottiene la (2.2).

Dalla (2.2) segue anche che

$$\lim_{x\to 0}\frac{1-\cos 2x}{x}=0\ . \tag{2.3}$$

Infatti, basta osservare che:

$$\frac{1-\cos 2x}{x}=\frac{1-\cos 2x}{x^2}\cdot x$$

e utilizzare ancora la *iii*) ed il limite (2.2).

I teoremi 2.3, 2.4 e 2.5 valgono anche se $x_0\in\dot{\mathbb{R}}$.

Le affermazioni *i*), ... , *v*) del teorema 2.5 non possono essere estese senza riserve al caso in cui i limiti l_1, l_2 che ivi compaiono siano elementi di $\mathbb{R}^*$ oppure di $\dot{\mathbb{R}}$ (cioè quando possono essere infiniti). Perdono poi significato quando uno o entrambi i limiti l_1, l_2 non esistono.

Valgono però molte affermazioni particolari, che qui sotto riportiamo sinteticamente. Le dimostrazioni di tali affermazioni sono tutte molto semplici e perciò vengono lasciate allo studente per esercizio.

■ **Teorema 2.6** - *Sia $x\to x_0$, con $x_0\in\mathbb{R}^*$ oppure $\dot{\mathbb{R}}$.*

i) $f(x)\to\infty$, $g(x)$ limitata (in un intorno di x_0) $\Rightarrow$ $(f\pm g)(x)\to\infty$.

Più precisamente si ha:

i_1) $f(x)\to+\infty$, $g(x)$ limitata inferiormente $\Rightarrow$ $(f+g)(x)\to+\infty$
i_2) $f(x)\to-\infty$, $g(x)$ limitata superiormente $\Rightarrow$ $(f+g)(x)\to-\infty$
i_3) $f(x)\to+\infty$, $g(x)$ limitata superiormente $\Rightarrow$ $(f-g)(x)\to+\infty$
i_4) $f(x)\to-\infty$, $g(x)$ limitata inferiormente $\Rightarrow$ $(f-g)(x)\to-\infty$.

Il teorema 2.6 autorizza ad adottare scritture del tipo: $(+\infty)+l=+\infty$, $(-\infty)+(-\infty)=-\infty$ e simili; queste possono anche semplificarsi (senza grave pregiudizio per la comprensione) scrivendo $+\infty+l=+\infty$, $-\infty-\infty=-\infty$ etc.

I casi mancanti all'elenco precedente conducono a espressioni del tipo

$$(+\infty)-(+\infty),\ (+\infty)+(-\infty),\ (-\infty)-(-\infty),\ (\infty)\pm(\infty)$$

(che possono compendiarsi nella forma $+\infty - \infty$) che sono *forme di indecisione*, poiché nessuna regola generale può fissarsi per esse. Ecco perché abbiamo parlato, a suo tempo, di una aritmetizzazione solo parziale dei simboli di infinito.

Esempi

2.13. $\lim_{x\to-\infty}(x+\sin x) = -\infty$; applicare la i_1.

2.14. $\lim_{x\to-\infty}(x^2 - x) = +\infty$; applicare la i_3.

2.15. $\lim_{x\to\infty}(x^2 + x)$: forma di indecisione (cfr. il seguente esempio 2.16).

■ **Teorema 2.7** - *Sia $x \to x_0$ ($x_0 \in \mathbb{R}^*$ oppure $\dot{\mathbb{R}}$);*

i) $f(x) \to \infty \;\Rightarrow\;$ la reciproca $1/f(x)$ è definita in un intorno di x_0 e si ha $1/f(x) \to 0$.

Più precisamente si ha:

i_1) $f(x) \to +\infty \;\Rightarrow\; 1/f(x) \to 0_+$
i_2) $f(x) \to -\infty \;\Rightarrow\; 1/f(x) \to 0_-$

ii) $f(x) \to 0$ e $f(x) \neq 0$ in un intorno di x_0, $x \neq x_0 \Rightarrow 1/f(x) \to \infty$.

Più precisamente si ha:

ii_1) $f(x) \to 0_+ \;\Rightarrow\; 1/f(x) \to +\infty$
ii_2) $f(x) \to 0_- \;\Rightarrow\; 1/f(x) \to -\infty$.

■ **Teorema 2.8** - *Sia $x \to x_0$ ($x_0 \in \mathbb{R}^*$ oppure $\dot{\mathbb{R}}$); sia $l \in \mathbb{R}^*$ oppure $\dot{\mathbb{R}}$.*

i) $f(x) \to \infty,\ g(x) \to l \neq 0 \;\Rightarrow\; f(x)g(x) \to \infty$.

Più precisamente si ha:

i_1) $f(x) \to +\infty,\ g(x) \to l < 0 \;\Rightarrow\; f(x)g(x) \to -\infty$
i_2) $f(x) \to +\infty,\ g(x) \to l > 0 \;\Rightarrow\; f(x)g(x) \to +\infty$
i_3) $f(x) \to -\infty,\ g(x) \to l < 0 \;\Rightarrow\; f(x)g(x) \to +\infty$
i_4) $f(x) \to -\infty,\ g(x) \to l > 0 \;\Rightarrow\; f(x)g(x) \to -\infty$

ii) $f(x) \to \infty,\ g(x)$ limitata, $|g(x)| \geq c > 0 \;\Rightarrow\; f(x)g(x) \to \infty$
iii) $f(x) \to 0,\ g(x)$ limitata $\;\Rightarrow\; f(x)g(x) \to 0$.

I teoremi 2.7 e 2.8 giustificano l'uso di scritture del tipo:

$$3 \cdot \infty = \infty \;\;,\;\; \frac{-2}{+\infty} = 0_- \;\;,\;\; (+\infty)\cdot(-\infty) = -\infty$$

e simili. I casi che non rientrano nei teoremi precedenti conducono a espressioni del tipo:

$$0 \cdot \infty \ , \ \frac{\infty}{\infty} \ , \ \frac{0}{0} \text{ e simili } \left(\frac{+\infty}{-\infty} \ \ldots \right)$$

che sono altrettante forme di indecisione.

Esempi

2.16. $\lim_{x \to \infty} (x^2 + x) = \lim_{x \to \infty} x(x+1) = \infty$ (per la *i*) del Teor. 2.8).

2.17. $\lim_{x \to 0_-} \dfrac{3x^2+1}{x} = -\infty$ (per la ii_2) del Teor. 2.7).

2.18. $\lim_{x \to 1} \dfrac{x^2-1}{x^2-2x+1}$: forma di indecisione $\dfrac{0}{0}$; ma essendo $\dfrac{x^2-1}{x^2-2x+1} = \dfrac{x^2-1}{(x-1)^2} = \dfrac{x+1}{x-1}$ il limite proposto risulta ∞.

2.19. $\lim_{x \to -\infty} \dfrac{2x^3-3x+7}{x^2+5}$: forma di indecisione $\dfrac{\infty}{\infty}$; ma essendo $\dfrac{2x^3-3x+7}{x^2+5} = \dfrac{x^3(2-3/x^2+7/x^3)}{x^2(1+5/x^2)} = x\dfrac{2-3/x^2+7/x^3}{1+5/x^2}$ il limite proposto risulta $-\infty$.

2.20. $\lim_{x \to \infty} \dfrac{\sin x}{x} = 0$ (per la *i*) del Teor. 2.7 e la *iii*) del Teor. 2.8).

2.21. $\lim_{x \to 0} x^\alpha \sin \dfrac{1}{x} = 0$ $(\forall \alpha > 0)$ (per la *iii*) del Teor. 2.8).

Le figure seguenti illustrano gli esempi 2.20 e 2.21.

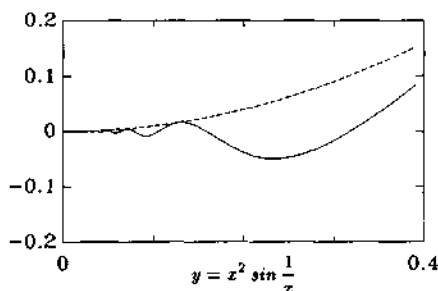
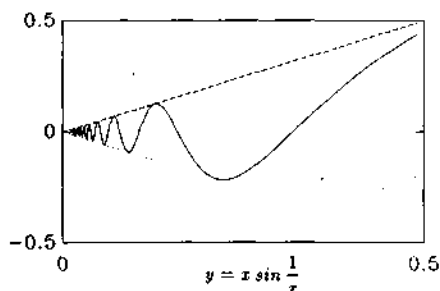
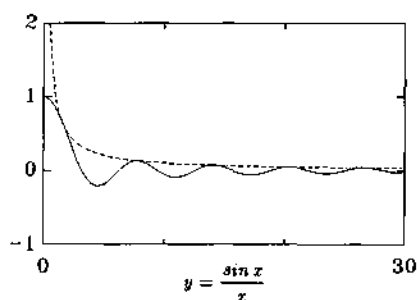



Fig. 4.10

Osservazione 2.5 - (Alcuni casi di non esistenza del limite)

Per comprendere meglio l'operazione di limite può essere utile esaminare qualche tipico caso in cui il limite, o meglio ancora il limite destro (o sinistro) non esiste.

Anzitutto dal teorema 2.2 sappiamo che il limite esiste se e solo se esistono i due limiti destro e sinistro e questi coincidono. Occupiamoci dell'esistenza del limite destro (considerazioni analoghe varranno per il limite sinistro).

Verifichiamo che la funzione: $f(x) = \sin 1/x$ non ammette limite per $x \to 0_+$.

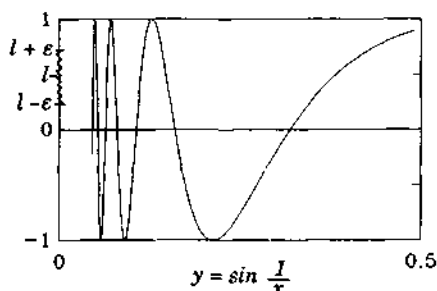



Fig. 4.11

Si noti anzitutto che: $-1 \leq \sin 1/x \leq 1$, cosicché, se il limite esistesse, dovrebbe appartenere all'intervallo $[-1,+1]$. Ma risulta $\sin(1/x) = 1$ nei punti $x_k = \dfrac{1}{\pi/2 + 2k\pi}$ $(k \in \mathbb{N})$ e $\sin(1/x) = -1$ nei punti $y_k = \dfrac{1}{3\pi/2 + 2k\pi}$ $(k \in \mathbb{N})$. Gli insiemi $\{x_k\}_{k\in\mathbb{N}}$ e $\{y_k\}_{k\in\mathbb{N}}$ hanno $x = 0$ come punto di accumulazione. Ne segue che in *ogni* intorno destro di 0 la funzione $\sin(1/x)$ assume infinite volte i valori $+1$ e -1. È chiaro perciò che non può esistere alcun intorno destro U_+ dell'origine tale che, se $0 \neq x \in U_+$, $f(x)$ resti confinata in un prefissato intervallo $(l-\varepsilon, l+\varepsilon)$ qualunque sia $l \in [-1,+1]$, se appena si sceglie $0 < \varepsilon < 1$.

Analogamente si dimostra (farlo per esercizio) che non esistono i seguenti limiti:

$$\lim_{x\to+\infty} x^\alpha \sin x \ \ (\alpha \geq 0) \qquad \lim_{x\to 0_+} x^\alpha \sin\frac{1}{x} \ \ (\alpha \leq 0)\,.$$

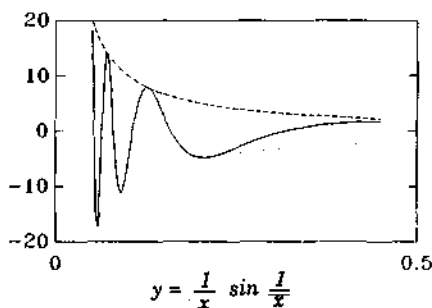
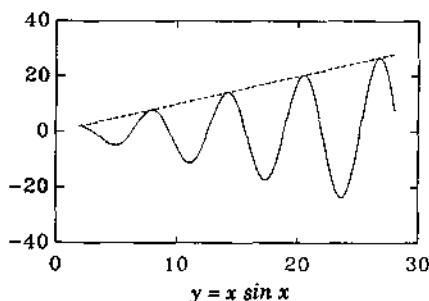



Fig. 4.12

Si confrontino i grafici riprodotti nella Fig. 10 (che illustrano casi di esistenza del limite) con i grafici delle Figg. 11 e 12 (che illustrano casi di non esistenza del limite). In tutti i casi la funzione $f(x)$ compie infinite oscillazioni quando x tende a x_0 ($x_0 = 0$ oppure $+\infty$) ma l'ampiezza di queste oscillazioni tende a zero nei primi casi, resta costante o addirittura si amplifica nei secondi.

Vediamo infine come si comporta l'operazione di limite rispetto al prodotto di composizione.

▪ **Teorema 2.9** - (Limite di funzione composta).

Sia $f : A \to \mathbb{R}$ *e* $g : B \to \mathbb{R}$ *con* $\operatorname{im}(g) \subseteq A$ *(cosicché è definita la composizione* $f \circ g : B \to \mathbb{R}$*). Sia* x_0 *un punto di accumulazione di* B *e risulti*

$$\lim_{x \to x_0} g(x) = l \tag{2.4}$$

con $g(x) \neq l$ *definitivamente per* $x \to x_0$,

$$\lim_{y \to l} f(y) = k \tag{2.5}$$

dove $x_0, l, k \in \mathbb{R}^*$ *(oppure* $\dot{\mathbb{R}}$*).*
Risulta allora

$$\lim_{x \to x_0} f(g(x)) = k \,. \tag{2.6}$$

Dimostrazione - Si osservi anzitutto che l è un punto di accumulazione per A, essendo, per la (2.4), un punto di accumulazione per $\operatorname{im}(g)$: ciò consente di formulare l'ipotesi (2.5). Assumiamo $x_0, l, k \in \mathbb{R}$. Per la (2.5), fissato $\varepsilon > 0$, esiste $\delta = \delta(\varepsilon)$ per cui: $|f(y) - k| < \varepsilon$ se $y \in A$ e $0 < |y - l| < \delta$. Per la (2.4), in corrispondenza di questo δ esiste $\sigma = \sigma(\delta)$ per cui: $0 < |g(x) - l| < \delta$ se $x \in B$ e $0 < |x - x_0| < \sigma$.
Perciò, fissato $\varepsilon > 0$, esiste $\sigma = \sigma(\varepsilon)$ per cui $|f(g(x)) - k| < \varepsilon$ se $x \in B$ e $0 < |x - x_0| < \sigma$, che prova la (2.6).
La precedente dimostrazione, condotta assumendo x_0, l, k in $\mathbb{R}$, si estende senza difficoltà al caso in cui qualcuno o anche tutti questi punti siano in $\mathbb{R}^*$ oppure $\dot{\mathbb{R}}$. □

Le (2.4), (2.5) e (2.6) del teorema 2.9 si possono interpretare come formule di *cambiamento di variabili* nel calcolo dei limiti.

***Esempio* 2.22** - Si voglia calcolare $\lim_{x \to \pm\infty} x \sin(1/x)$. Posto $y = 1/x$ si osserva che $\lim_{x \to \pm\infty} y = 0_\pm$.
Dunque $\lim_{x \to \pm\infty} x \sin(1/x) = \lim_{y \to 0_\pm} \sin y / y = 1$ (cfr. esempio 2.11).

I teoremi 2.5–2.9 indicano come operare per il calcolo dei limiti tranne che nei casi che abbiamo battezzato *forme di indecisione*. Come ci si deve comportare in presenza di una forma di indecisione? Come si può decidere se il limite esiste ed eventualmente calcolarlo? Qualche "trucco" è indicato negli esempi 2.16–2.19.

Vedremo più avanti due strumenti estremamente potenti (teorema di de l'Hôpital e formula di Taylor) che permetteranno di operare agevolmente nella quasi totalità dei casi di indecisione.

Tuttavia il punto chiave nella risoluzione di una forma di indecisione sta nel valutare le "diverse velocità" con cui una funzione tende a zero o all'infinito. Ciò è oggetto del par. 2.5.

2.4 Esistenza del limite (per funzioni monotone)

Ci proponiamo ora di trovare delle condizioni sufficienti a garantire la possibilità di eseguire l'operazione di limite. A questo proposito la classe delle funzioni monotone merita una considerazione particolare poiché per tali funzioni questa operazione risulta sempre possibile.

■ **Teorema 2.10** - *Siano a, b elementi di $\mathbb{R}^*, a < b$.*

i) Sia $f : (a,b) \to \mathbb{R}$ una funzione crescente.

i_1) Per ogni $x_0 \in (a,b]$ si ha: $\lim_{x\to x_0-} f(x) = \sup_{(a,x_0)} f$.

i_2) Per ogni $x_0 \in [a,b)$ si ha: $\lim_{x\to x_0+} f(x) = \inf_{(x_0,b)} f$.

ii) Sia $f:(a,b) \to \mathbb{R}$ una funzione decrescente.

ii_1) Per ogni $x_0 \in (a,b]$ si ha: $\lim_{x\to x_0-} f(x) = \inf_{(a,x_0)} f$.

ii_2) Per ogni $x_0 \in [a,b)$ si ha: $\lim_{x\to x_0+} f(x) = \sup_{(x_0,b)} f$.

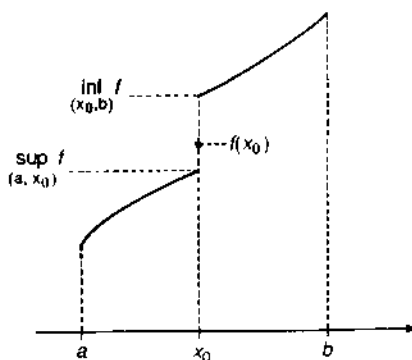



Fig. 4.13 Esistenza del limite per una funzione crescente:

$$\lim_{x\to x_{0-}} f = \sup_{(a,x_0)} f \quad \text{e} \quad \lim_{x\to x_{0+}} f = \inf_{(x_0,b)} f$$

Si noti che $f(x_0)$ è diversa dai due limiti.

Dimostrazione - Dimostriamo soltanto i_1; le altre dimostrazioni sono analoghe e vengono lasciate per esercizio. Sia $\sup\limits_{(a,x_0)} f = l$ e distinguiamo due casi: l finito, $l = +\infty$.

Se l è finito, per le proprietà dell'estremo superiore di f (cfr. 1.2) risulta

$$f(x) \leq l \qquad \forall x \in (a, x_0)$$

$$\forall \varepsilon > 0, \quad \exists x_\varepsilon \in (a, x_0)$$

tale che: $l - \varepsilon < f(x_\varepsilon)$ sicché, essendo f crescente, si ha:

$$l - \varepsilon < f(x) \leq l \qquad \forall x \in (x_\varepsilon, x_0)\ ;$$

ciò prova che $\lim\limits_{x \to x_{0-}} f(x) = l$.

Se poi $l = +\infty$ (e in tal caso non può che essere $x_0 = b$, perché se fosse $x_0 < b$ sarebbe, essendo f crescente, $\sup\limits_{(a,x_0)} f \leq f(x_0)$) abbiamo:

$$\forall k > 0, \quad \exists x_k \in (a, b)$$

tale che:

$$f(x_k) > k\ ,$$

sicché risulta

$$f(x) > k \qquad \forall x \in (x_k, b)\ ;$$

ciò prova che $\lim\limits_{x \to b_-} f(x) = +\infty$. □

Esempi (Potenze, esponenziali, logaritmi)

2.23. Sia $\alpha \neq 0, x_0 \in \mathbb{R}_+$. Valgono le relazioni

$$\lim_{x \to x_0} x^\alpha = x_0^\alpha \quad \lim_{x \to 0_+} x^\alpha = \begin{cases} 0 & (\alpha > 0) \\ +\infty & (\alpha < 0) \end{cases} \quad \lim_{x \to +\infty} x^\alpha = \begin{cases} +\infty & (\alpha > 0) \\ 0 & (\alpha < 0) \end{cases}$$

Esse possono ricavarsi dalla monotonia delle funzioni $x \mapsto x^\alpha$ (definite su $\mathbb{R}_+$) e dalle proprietà delle potenze (cfr. 2.3.2).

Per provare la prima conviene scrivere $x^\alpha = (x/x_0)^\alpha x_0^\alpha$ e, posto $t = x/x_0$, mostrare che $\lim\limits_{t \to 1} t^\alpha = 1$ (farlo per esercizio).

2.24. Sia $a > 0, a \neq 1, x_0 \in \mathbb{R}$. Valgono le relazioni

$$\lim_{x \to x_0} a^x = a^{x_0} \quad \lim_{x \to +\infty} a^x = \begin{cases} +\infty & (a > 1) \\ 0_+ & (a < 1) \end{cases} \quad \lim_{x \to -\infty} a^x = \begin{cases} 0_+ & (a > 1) \\ +\infty & (a < 1) \end{cases}$$

che derivano dalla monotonia delle funzioni esponenziali e dalle altre loro proprietà (cfr. 2.3.2). Ancora, per provare la prima, conviene scrivere $a^x = a^{x_0} \cdot a^{x - x_0}$, porre $t = x - x_0$, e provare che $\lim\limits_{t \to 0} a^t = 1$ (farlo per esercizio).

2.25. Sia $a > 0, a \neq 1$, $x_0 \in \mathbb{R}_+$. Valgono le relazioni

$$\lim_{x \to x_0} \log_a x = \log_a x_0 \qquad \lim_{x \to 0_+} \log_a x = \begin{cases} -\infty & (a > 1) \\ +\infty & (a < 1) \end{cases}$$

$$\lim_{x \to +\infty} \log_a x = \begin{cases} +\infty & (a > 1) \\ -\infty & (a < 1) \end{cases}$$

che possono essere dedotte dalle precedenti, ricordando che il logaritmo è la funzione inversa dell'esponenziale, e utilizzando il teorema 2.9.

Osservazione 2.6 - Utilizzando i risultati degli esempi 2.24 e 2.25 e il teorema 2.9 possiamo semplificare il calcolo di limiti che si presentano sotto la forma

$$\lim_{x \to x_0} f(x)^{g(x)} \tag{2.7}$$

(con $f(x)$ definitivamente positiva).

Infatti possiamo scrivere, scegliendo una base $a \neq 1$,

$$f(x)^{g(x)} = a^{g(x) \log_a f(x)}$$

e perciò il limite (2.7) viene ricondotto a uno del tipo

$$\lim_{x \to x_0} g(x) \log_a f(x) \ . \tag{2.8}$$

Si vede così che, accanto alle forme di indecisione già viste ($\infty - \infty, 0/0, \infty/\infty$, $0 \cdot \infty$) sono da considerarsi forme di indecisione anche le seguenti

$$0^0 \quad , \quad 1^\infty \quad , \quad \infty^0$$

che, per quanto detto, si possono però ricondurre alla forma di indecisione $0 \cdot \infty$ (si noti che 0^∞ non è una forma di indecisione!)

***Esempio* 2.26** - Si voglia calcolare $\lim_{x \to 0_+} x^x$ e $\lim_{x \to +\infty} x^x$.

Basterà calcolare (con $a > 1$) $\lim_{x \to 0_+} x \log_a x$ e $\lim_{x \to +\infty} x \log_a x$. Il primo è una forma di indecisione $0 \cdot \infty$ che sarà risolto in 3.2 e risulta 0 (cfr. esercizio 3.12); il secondo è $+\infty$. Perciò abbiamo

$$\lim_{x \to 0_+} x^x = 1 \quad , \quad \lim_{x \to +\infty} x^x = +\infty \ .$$

Esercizi

7. Dimostrare che se $f(x) > 0$ definitivamente per $x \to x_0$ e se esiste $\lim_{x \to x_0} f(x) = l$ allora $l \geq 0$.

8. Calcolare mediante il teorema del confronto

a) $\lim\limits_{x\to+\infty} x(a+\sin x) \quad (a\in\mathbb{R})$

b) $\lim\limits_{x\to+\infty} \dfrac{1}{(2-\cos x)x^{\alpha}} \quad \alpha\in\mathbb{R}_+$.

9. Siano

$$f(x)=\begin{cases}0 & \text{se } x=0\\ 1 & \text{se } x\neq 0\end{cases} \quad \text{e} \quad g(x)=x\sin\frac{1}{x}.$$

Si ha: $f(x)\to 1$ e $g(x)\to 0$ per $x\to 0$. Si può applicare il teorema 2.9? Giustificare la risposta.

10. Calcolare i seguenti limiti

a) $\lim\limits_{x\to 0} \dfrac{\operatorname{tg} 3x}{\sin 2x}$

b) $\lim\limits_{x\to 1} \dfrac{(x-1)^2}{\sin \pi x}$

c) $\lim\limits_{x\to+\infty} x^{\alpha}\sin\dfrac{1}{x}\ (\alpha>0)$.

11. Siano:

$$P(x)=a_n x^n + a_{n-1}x^{n-1}+\cdots+a_1 x + a_0$$

e

$$Q(x)=b_m x^m + b_{m-1}x^{m-1}+\cdots+b_1 x + b_0 ,$$

(a_n e b_m non nulli).
Discutere al variare di n ed m,

$$\lim_{x\to\pm\infty}\frac{P(x)}{Q(x)} .$$

12. Sia $f(x)=\mathbf{1}_{\mathbb{Q}}(x)$ la funzione caratteristica dei razionali. Verificare che $\forall x_0\in\mathbb{R}^*$ (o $\mathbb{R}$), $\lim\limits_{x\to x_0} f(x)$ non esiste.

13. Dimostrare che $\lim\limits_{x\to+\infty} \operatorname{tg} x$ non esiste.

14. Calcolare i seguenti limiti

a) $\lim\limits_{x\to 0_\pm} 2^{1/x}$

b) $\lim\limits_{x\to 0_+} (\sin x)^{1/(\cos x-1)}$.

c) $\lim\limits_{x\to \pi_-} (\operatorname{tg} x)^{1/(\pi-x)}$. (!)

15. Calcolare i seguenti limiti

a) $\lim_{x\to\pm\infty}(\sqrt{x^4-x^2}-x^2)$

b) $\lim_{x\to\pm\infty}(\sqrt[3]{x^3-x}-x)$.

2.5 Infinitesimi ed infiniti. Confronti.

Siano f e g due funzioni definite in un intorno di $x_0 \in \mathbb{R}^*$, tranne al più nel punto x_0 stesso.

Si dice **infinitesimo** per $x \to x_0$ ogni funzione $f(x)$ che tende a zero per $x \to x_0$. Si dice **infinito** per $x \to x_0$ ogni funzione che tende a $+\infty$ oppure a $-\infty$ (o anche a ∞).

Esempi

2.27. $\sin x, x^{1/3}, \log_2(1+x)$ sono infinitesimi per $x \to 0$; $1/\sqrt{x}$ e 2^{-x} sono infinitesimi per $x \to +\infty$; $\log(1-x)$ è infinito per $x \to 1$.

Quando le due funzioni f e g sono entrambe *infinitesimi* per $x \to x_0$ è utile poter stabilire un *confronto* tra esse. A tale scopo, se g è definitivamente $\neq 0$ per $x \to x_0$, consideriamo il limite di f/g; si hanno quattro possibilità:

$$\lim_{x\to x_0}\frac{f(x)}{g(x)} = \begin{array}{ll} 1. & 0 \\ 2. & l \text{ finito e diverso da zero} \\ 3. & \text{infinito } (+\infty, -\infty \text{ o } \infty) \\ 4. & \text{non esiste.} \end{array}$$

Diciamo che, se vale

1: f è infinitesimo di *ordine superiore* rispetto a g;

2: f è dello *stesso ordine* di g;

3: f è infinitesimo di *ordine inferiore* rispetto a g;

4: f e g *non* sono confrontabili.

Se f e g sono entrambe *infiniti* per $x \to x_0$ in corrispondenza alle quattro possibilità indicate sopra per il limite del loro rapporto, diremo che:

se vale:

1: f è infinito di *ordine inferiore* rispetto a g;

2: f è dello *stesso ordine* di g;

3: f è infinito di *ordine superiore* rispetto a g;

4: f e g *non* sono confrontabili.

Esempi

2.28. x e $\sin x$ sono infinitesimi dello stesso ordine per $x \to 0$ (cfr. esempio 2.11); così pure x^2 e $1 - \cos 2x$ (cfr. esempio 2.12).

2.29. Per $x \to +\infty$ le funzioni x^α $(\alpha > 0)$, a^x $(a > 1)$ e $\log_a x$ sono infiniti (cfr. esempi 2.23 – 2.25).

Il confronto tra questi infiniti sarà fatto nel paragrafo 3.2.

Se f_1 è infinitesimo di ordine superiore a f (per $x \to x_0$) diremo che f_1 è *trascurabile* rispetto a f o anche che f è *dominante* rispetto a f_1; questo modo di dire è giustificato dal fatto che, se f, f_1, g, g_1 sono infinitesimi, f_1 di ordine superiore ad f, g_1 di ordine superiore a g, allora risulta

$$\lim_{x \to x_0} \frac{f + f_1}{g + g_1} = \lim_{x \to x_0} \frac{f}{g} \frac{1 + f_1/f}{1 + g_1/g} = \lim_{x \to x_0} \frac{f}{g}$$

cioè *sia a numeratore che a denominatore si possono di fatto omettere, per il calcolo del limite, gli infinitesimi di ordine superiore.*

***Esempio* 2.30** - $\displaystyle \lim_{x \to 0} \frac{3x + (\sin x)^2 + 2x^4}{x^3 - x} = \lim_{x \to 0} \frac{3x}{-x} = -3 .$

Analogamente, se f_1 è infinito di ordine inferiore a f diremo che f_1 è trascurabile rispetto a f.

***Esempio* 2.31** - $\displaystyle \lim_{x \to -\infty} \frac{x^2 + 1 + |x|^{1/2}}{3x - x^3} = \lim_{x \to -\infty} \frac{x^2}{-x^3} = \lim_{x \to -\infty} -\frac{1}{x} = 0 .$

Sono entrati nell'uso corrente alcuni simboli (detti simboli di Landau), utili nell'esprimere relazioni di confronto tra funzioni.

1. Il simbolo $o(\cdot)$ ("o piccolo di $\cdots$")

Se g è definitivamente $\neq 0$ per $x \to x_0$, allora

$$f(x) = o(g(x)) \quad \text{significa} \quad \lim_{x \to x_0} \frac{f(x)}{g(x)} = 0$$

ovvero che f è infinitesimo di ordine superiore a g per $x \to x_0$. Osserviamo che $f(x) = o(1)$ significa semplicemente che $f(x)$ è infinitesimo. Perciò, la formula

$$\lim_{x \to x_0} f(x) = l, \quad \text{se } l \in \mathbb{R}, \text{ equivale a } \quad f(x) = l + o(1) \quad (x \to x_0) .$$

2. Il simbolo $O(\cdot)$ ("O grande di $\cdots$")

Se $g(x)$ è definitivamente $\neq 0$ per $x \to x_0$, allora

$$f(x) = O(g(x)) \quad \text{significa} \quad \frac{f(x)}{g(x)} \quad \text{definitivamente limitato per } x \to x_0.$$

In particolare, $f(x) = O(1)$ significa che $f(x)$ è definitivamente limitata per $x \to x_0$ (ciò non esclude che $f(x)$ possa essere infinitesima).

3. Il simbolo $\sim$ ("asintotico a $\cdots$")

Se $g(x)$ è definitivamente $\neq 0$ per $x \to x_0$, allora

$$f(x) \sim g(x) \quad \text{significa} \quad \lim_{x \to x_0} \frac{f(x)}{g(x)} = 1 .$$

Una semplice ed importante proprietà del simbolo di asintotico è la seguente:

se $f \sim f_1$, e $g \sim g_1$ per $x \to x_0$ allora $f/g \sim f_1/g_1$ che in molti casi risulta utile nel calcolo dei limiti.

Si osserverà che nella classe delle funzioni definitivamente diverse da zero per $x \to x_0$ la relazione $\sim$ è riflessiva, simmetrica, transitiva; tali funzioni si distribuiscono perciò in classi di funzioni a due a due asintotiche.

4. Il simbolo $\asymp$ ("stesso ordine di grandezza di $\cdots$")

Se $g(x)$ è definitivamente $\neq 0$ per $x \to x_0$, allora $f(x) \asymp g(x)$ significa che esistono due numeri positivi, m e M, $0 < m < M$, tali che si abbia definitivamente per $x \to x_0$,

$$m|g(x)| \leq |f(x)| \leq M|g(x)| .$$

In particolare, $f(x) \asymp 1$ significa che sia $f(x)$ che $1/f(x)$ si mantengono limitate in un intorno di x_0.

Anche la relazione binaria $\asymp$ è un'equivalenza nella classe delle funzioni definitivamente $\neq 0$ per $x \to x_0$.

Esempi

2.32. Sia $x \to 0$. Verifichiamo le relazioni seguenti.

a) $\sin x \sim x$

b) $\operatorname{tg} x \sim x$

c) $\sin x = o(1)$

d) $\operatorname{tg} x = O(x)$

e) $x = o(\sqrt{|x|})$

f) $3x^2 + x \asymp x$

g) $x^2 + 3x^4 = O(x^2)$.

a) Dall'esempio 2.11 $\dfrac{\sin x}{x} \to 1$

b) $\dfrac{\operatorname{tg} x}{x} = \dfrac{\sin x}{x} \dfrac{1}{\cos x} \to 1$

c) Equivale a $\sin x \to 0$

d) Essendo $\frac{\operatorname{tg} x}{x} \to 1$, $\frac{\operatorname{tg} x}{x}$ è definitivamente limitata per $x \to 0$

e) $\frac{x}{\sqrt{|x|}} = \operatorname{sign} x \cdot \sqrt{|x|} \to 0$

f) $\frac{3x^2 + x}{x} = 3x + 1 \to 1$ e quindi $3x + 1$ e $\frac{1}{3x+1}$ sono definitivamente limitate per $x \to 0$

g) $\frac{x^2 + 3x^4}{x^2} = 1 + 3x^2 \to 1$ e quindi $1 + 3x^2$ è definitivamente limitata.

L'allievo verifichi che sono false le relazioni

$$1 - \cos x \sim x^2 \ , \ \sin x = o(\operatorname{tg} x) \ , \ x^2 = O(x^3) \ .$$

2.33. Sia $x \to +\infty$. Sono vere le relazioni

$$\cos x = O(1), \ \log_a x = O(x) \ (a > 1)$$

$$[x] \sim x, \ \sqrt{x^2 - 1} \sim x, \ x + \sqrt{x^2 - 1} \asymp x \ .$$

Sono false le relazioni

$$x + \sqrt{x^2 - 1} \sim x, \ \cos x = o(1), \ x^3 \sim x^2 \ .$$

Asintoti.

Interessante è il caso

$$f(x) = ax + b + o(1) \ \text{ per } x \to +\infty \ (\text{oppure } -\infty). \tag{2.9}$$

La (2.9) esprime il fatto che f si "comporta" come un polinomio di primo grado quando $x \to +\infty$ $(-\infty)$.

In tal caso la retta $y = ax + b$ viene detta *asintoto* per la funzione f. Se $a = 0$ si ha un *asintoto orizzontale, obliquo* se $a \neq 0$.

L'interpretazione geometrica della (2.9) è la seguente. Indichiamo con r la retta di equazione $y = ax + b$ e con P_x il punto di coordinate $(x, f(x))$.

Se $d(P_x, r)$ indica la distanza di P_x da r si ha (*):

$$d(P_x, r) = \frac{|ax + b - f(x)|}{\sqrt{a^2 + 1}} \ .$$

Dunque la (2.9) equivale a $d(P_x, r) \to 0$ per $x \to +\infty$ $(-\infty)$.

(*) La formula per la distanza di un punto $P = (x_0, y_0)$ dalla retta di equazione $mx + ny + q = 0$ è $|mx_0 + ny_0 + q|/\sqrt{m^2 + n^2}$.

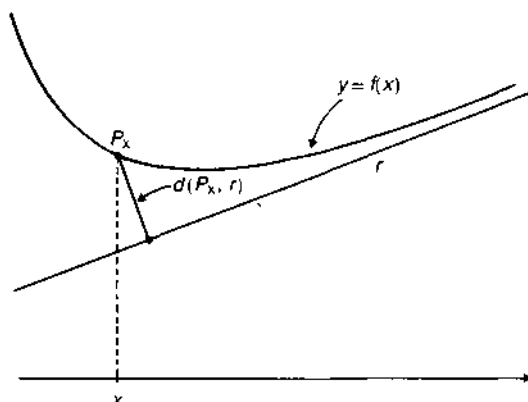



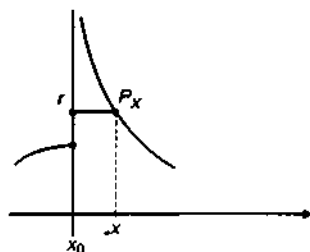
Fig. 4.14 La retta r è asintoto obliquo per f quando $x \to +\infty$.

È facile poi verificare che la (2.9) è equivalente alle seguenti due condizioni, a volte di più facile uso nella determinazione di eventuali asintoti:

$$(2.9) \Leftrightarrow \frac{f(x)}{x} \to a \in \mathbb{R}, \text{ e } f(x) - ax \to b \in \mathbb{R} \ . \tag{2.10}$$

Accanto agli asintoti orizzontali ed obliqui, vi sono quelli verticali:

se $x_0 \in \mathbb{R}$ e almeno uno dei limiti per $x \to x_{0+}$ o $x \to x_{0-}$ di f è $\pm\infty$ la retta di equazione $x = x_0$ si dice *asintoto verticale* per f.

Fig. 4.15 Poiché $\lim_{x \to x_{0+}} f(x) = +\infty$, la retta $x = x_0$ è asintoto verticale (da destra) per f. Si vede che $d(P_x, r) \to 0$ per $x \to x_{0+}$.

Esempi

2.33. a) Si ha: $\lim_{x \to +\infty} \frac{x}{x+1} = 1$; dunque $y = 1$ è asintoto orizzontale per $f(x) = \frac{x}{x+1}$.

b) $\lim_{x \to -1_\pm} \frac{x}{x+1} = \mp\infty$; dunque $x = -1$ è asintoto verticale (da destra e da sinistra) per $f(x) = \frac{x}{x+1}$.

c) Sia $f(x) = \sqrt{x^2+1}$. Poiché $\dfrac{\sqrt{x^2+1}}{x} = \dfrac{|x|\sqrt{1+1/x^2}}{x}$, si ha
$\lim_{x \to \pm\infty} \dfrac{\sqrt{x^2+1}}{x} = \pm 1$.

Essendo poi $\sqrt{x^2+1} - x = \dfrac{1}{\sqrt{x^2+1}+x} \to 0$ per $x \to +\infty$ e $\sqrt{x^2+1} + x =$
$= \dfrac{1}{\sqrt{x^2+1}-x} \to 0$ per $x \to -\infty$, concludiamo, in base alla (2.10), che $y = x$ è asintoto obliquo per $x \to +\infty$ mentre $y = -x$ lo è per $x \to -\infty$.

d) Sia $f(x) = x + \sqrt[3]{x}$. Per $x \to +\infty$ $f(x) \sim x$ ma $y = x$ non è asintoto obliquo per f in quanto $f(x) - x \not\to 0$.

Il confronto tra infinitesimi e infiniti può essere a volte reso più preciso "misurando la velocità" con cui una funzione tende a zero o all'infinito rispetto ad un infinitesimo o ad un infinito, scelto come *campione* (una specie di *unità di misura*).

Ci riferiremo agli infinitesimi, ma tutto si può estendere con facilità al caso degli infiniti.

Sia g un infinitesimo positivo per $x \to x_0$ che scegliamo come *campione*.

Se $f(x)$ è un altro infinitesimo (per $x \to x_0$) ed esistono due numeri reali α e k, diversi da zero, tali che :

$$f(x) \sim k g(x)^\alpha$$

allora si dice che f è *di ordine α rispetto* a g e che $k\,g(x)^\alpha$ è la *parte principale* dell'infinitesimo $f(x)$.

Esempi

2.35. Sia $x \to 0_+$ e $g(x) = x$ l'infinitesimo campione.

a) $\forall \alpha \in \mathbb{R}_+$ $\quad f(x) = x^\alpha$ è di ordine α rispetto ad x (ovvio!)

b) $\forall \alpha \in \mathbb{R}_+$ $\quad \sin(x^\alpha)$ è di ordine α rispetto ad x. Infatti posto $x = t^\alpha$ si ha $\sin t \sim t$ e quindi $\sin(x^\alpha) \sim x^\alpha$.

c) $3x^5 + 6x^3 + x^2$ è di ordine 2 rispetto ad x. Infatti $3x^5 + 6x^3 + x^2 \sim x^2$.

2.36. Sia $x \to +\infty$ e $g(x) = 1/x$ l'infinitesimo campione. Allora:

a) $\sin \dfrac{1}{\sqrt{x}}$ è di ordine $1/2$

b) $\dfrac{x}{x^3+1}$ è di ordine 2.

c) $\dfrac{\sqrt{x}+1}{x^3-x}$ è di ordine $5/2$.

Lasciamo al lettore la verifica di queste affermazioni.

Esercizi

16. Vero o falso:

a) $\dfrac{\sin x}{x} \sim \dfrac{1}{x}$ per $x \to +\infty$

b) $\operatorname{tg}\left(\dfrac{1}{\sqrt{x}}\right) \sim \dfrac{1}{\sqrt{x}}$ per $x \to +\infty$

c) $\dfrac{\sin x^\alpha}{\sin x^\beta} \sim x^{\alpha-\beta}$ per $x \to 0_+$ $(\alpha, \beta \in \mathbb{R}_+)$

d) $\dfrac{x^\alpha + \sin x}{x^\beta - \cos x} \sim x^{\alpha-\beta}$ per $x \to +\infty$ $(\alpha, \beta \in \mathbb{R}_+)$.

17. Dimostrare, esibendo un controesempio, che $f \sim f_1$ e $g \sim g_1$ *non* implica $(f-g) \sim (f_1 - g_1)$.

18. Sia $x \to 0_+$. Dimostrare le seguenti proprietà del simbolo o:

a) $\displaystyle\sum_{j=1}^{k} o(x^{\alpha_j}) = o(x^{\min \alpha_j})$ $(\alpha_j \in \mathbb{R}_+$ per ogni $j = 1, \cdots, k)$

b) $x^\alpha \cdot o(x^\beta) = o(x^{\alpha+\beta})$ e $o(x^\alpha) \cdot o(x^\beta) = o(x^{\alpha+\beta})$

c) $o(o(x^\alpha)) = o(x^\alpha)$.

19. Vero o falso. Sia $x \to +\infty$.

a) $2 + \sin x = O(1)$

b) $x \operatorname{tg} x \asymp x$

c) $\sqrt{x^2 - x} - x \asymp 1$

d) $\dfrac{\sqrt{x^2+1} + \cos x}{x^3 + x + 1} = o\left(\dfrac{1}{x}\right)$.

20. Scriviamo "$\sim \Rightarrow \asymp$" come abbreviazione di $f \sim g \Rightarrow f \asymp g$. Analogamente per le altre coppie di simboli.

Completare la tabella seguente riguardante le mutue implicazioni tra i simboli di Landau: gli antecedenti sono riportati nella colonna a sinistra (Inserire V quando l'implicazione è ritenuta vera e F quando falsa).

	$\sim$	o	$\asymp$	O
$\sim\Rightarrow$				
$o \Rightarrow$				
$\asymp\Rightarrow$				
$O \Rightarrow$				

21. Vero o falso: siano $f, f_1 \to +\infty$

a) $f_1 \sim f \Rightarrow f_1 - f \to 0$
b) $f_1 - f \to 0 \Rightarrow f_1 \sim f$.

22. Vero o falso: siano $f, f_1 \to +\infty$, $a \in \mathbb{R}_+, a \neq 1$.

a) $f \sim f_1 \Rightarrow a^f \sim a^{f_1}$
b) $f \sim f_1 \Rightarrow a^f \asymp a^{f_1}$
c) $a^f \sim a^{f_1} \Rightarrow f \sim f_1$
d) $f \sim f_1 \Rightarrow \log_a f \sim \log_a f_1$.

23. È vero che se $f(x) \sim ax + b$ per $x \to +\infty$ la retta $y = ax + b$ è asintoto obliquo per f?

24. Per quale/i tra le seguenti funzioni si può escludere (senza calcoli) l'esistenza di un *asintoto obliquo per* $x \to +\infty$?
Per le altre calcolare l'eventuale asintoto obliquo.

a) $\log_a(x^3 + 1)$
b) $x(\sin x + 3)$
c) $\dfrac{x^3 + 1}{x^2 + x + 1}$.
d) $\dfrac{2^x}{x}$.

25. Determinare l'ordine di infinitesimo rispetto a $g(x) = x$, per $x \to 0$ delle seguenti funzioni e la loro parte principale.

a) $\cos x - 1$
b) $(\sin 3\sqrt[3]{x})^2$
c) $\sqrt[3]{\operatorname{tg} x}$
d) $\dfrac{1}{\sqrt[3]{x}}(1 - (\cos x)^3)$.

3. SUCCESSIONI A VALORI IN $\mathbb{R}$

3.1 Limite di una successione

Una classe di funzioni che merita una trattazione particolare per quel che riguarda l'operazione di limite è costituita dalle successioni, cioè dalle funzioni definite su $\mathbb{N}$. La situazione risulta semplificata, rispetto al caso generale delle funzioni, dal momento che l'unico punto di accumulazione di $\mathbb{N}$ è $+\infty$, e perciò si presentano

solo i seguenti casi: sia $\{a_n\}$ una successione con valori in $\mathbb{R}$ (*):

i) $\lim_{n\to+\infty} a_n = l \in \mathbb{R}$

ii) $\lim_{n\to+\infty} a_n = +\infty, -\infty, \infty$

iii) $\lim_{n\to+\infty} a_n$ non esiste.

Nel primo caso la successione si dirà *convergente*, nel secondo *divergente*, e nel terzo *irregolare* (o *indeterminata*). Per comodità dell'allievo ripetiamo ancora la definizione di limite adattandola alla situazione presente:

i) $\lim_{n\to+\infty} a_n = l \in \mathbb{R}$ significa che: fissato $\varepsilon > 0$, in modo arbitrario, esiste un N (dipendente da questo $\varepsilon : N = N(\varepsilon)$) tale che: se $n \geq N$ risulta $|a_n - l| < \varepsilon$.

ii) $\lim_{n\to+\infty} a_n = -\infty$ significa che: fissato $k > 0$, in modo arbitrario, esiste un $N = N(k)$ tale che: se $n \geq N$ risulta $a_n < -k$.

(Ripetere la definizione quando il limite è $+\infty$ oppure ∞).

Le analoghe definizioni per il limite per eccesso o per difetto si adattano senza difficoltà.

Dal calcolo dei limiti sviluppato in 2.3 si ricavano immediatamente le seguenti proposizioni.

Proposizioni

3.1. (Permanenza del segno) *Se $a_n \to l > 0$ (< 0) allora è definitivamente* $a_n > 0$ (< 0) (non vale il viceversa) (cfr. teor. 2.3).

Si osservi che la locuzione "*definitivamente*" per le successioni significa: "per n abbastanza grande", cioè per n maggiore di un opportuno N. Per esempio, la successione $n - 10\sqrt{n}$ è definitivamente positiva: infatti è positiva per $n > 100$.

3.2. *Se $\{a_n\}$ è convergente, allora è limitata* (non vale il viceversa) (cfr. osserv. 2.2).

3.3. (Confronto) *Se è definitivamente $a_n \leq b_n \leq c_n$ e se $a_n \to l$, $c_n \to l$, allora anche $b_n \to l$* (cfr. teor. 2.4).

3.4. *Se è definitivamente $a_n \leq b_n$ e se $a_n \to +\infty$, allora anche $b_n \to +\infty$;* analoga proposizione *se $b_n \leq c_n$ e $c_n \to -\infty$.*

I teoremi 2.5, 2.6, 2.7, 2.8 del calcolo dei limiti si semplificano (per la minor varietà dei casi in esame). Vale anche il teorema 2.9, opportunamente adattato. Il teorema 2.10 sui limiti delle successioni monotone si può enunciare così:

(*) Anche se non sarà ogni volta detto esplicitamente, tutte le successioni considerate in questa sezione sono a valori in $\mathbb{R}$.

3.5. *Una successione (definitivamente) crescente ammette sempre limite* $l = \sup\{a_n\}$; *tale limite è perciò finito se* $\{a_n\}$ *è limitata superiormente, altrimenti* $l = +\infty$. Analogo è l'enunciato per le successioni decrescenti.

Esempi

3.1. $\frac{1}{n} \to 0$; infatti, fissato $\varepsilon > 0$, risulta $\frac{1}{n} < \varepsilon$ per $n > \frac{1}{\varepsilon}$; basterà perciò prendere $N = [1/\varepsilon] + 1$. Può essere interessante notare che per questa semplice dimostrazione è necessario utilizzare il principio di Archimede!

3.2. $(-1)^n$ è irregolare.

3.3. $\frac{n}{n+1} \to 1_-$; basta osservare che $\frac{n}{n+1} = 1 - \frac{1}{n+1}$.

3.4. $\sqrt[n]{a} \to 1$ $(a > 0)$; sia $a > 1$ e poniamo $\sqrt[n]{a} = 1 + h_n$ (con $h_n > 0$); risulta $a = (1 + h_n)^n \geq 1 + nh_n$ per la (3.11) del cap. 2 (disuguaglianza di Bernoulli) da cui si ricava che $0 < h_n \leq (a-1)/n$; applicando il criterio del confronto (prop. 3.3) si deduce che $h_n \to 0$.
Se è $a < 1$ si scriverà $\sqrt[n]{a} = 1/\sqrt[n]{b}$ (con $b = 1/a > 1$) e applicheremo il risultato precedente insieme col teorema 2.5 (iv) del calcolo dei limiti.

3.5. *La progressione geometrica.*

$$\lim_{n \to +\infty} q^n = \begin{cases} 0 & \text{se } |q| < 1 \\ 1 & \text{se } q = 1 \\ +\infty & \text{se } q > 1 \\ \text{non esiste} & \text{se } q = -1 \\ \infty & \text{se } q < -1 \text{ (non esiste in } \mathbb{R}^*\text{; esiste in } \dot{\mathbb{R}}) . \end{cases}$$

Nel primo caso, posto $q = \frac{1}{1+p}$ (con $p > 0$) risulta

$$0 < q^n = \frac{1}{(1+p)^n} \leq \frac{1}{1+np} < \frac{1}{np} \to 0 .$$

Nel terzo caso, posto $q = 1 + p$ (con $p > 0$) risulta

$$q^n = (1+p)^n \geq 1 + np \to +\infty .$$

Nel quinto caso, i termini della successione sono alternativamente positivi e negativi; mentre in valore assoluto tendono a $+\infty$; perciò il limite non esiste in $\mathbb{R}^*$, ma esiste in $\dot{\mathbb{R}}$ ed è ∞.

3.6. $\lim_{n \to +\infty} \sqrt{n+1} - \sqrt{n-1}$; per risolvere la forma di indecisione $\infty - \infty$ basterà scrivere: $\sqrt{n+1} - \sqrt{n-1} = \frac{2}{\sqrt{n+1} + \sqrt{n-1}}$; il limite proposto è perciò 0.

Osservazione 3.1 - Se $\{a_n\}$ è una *successione stabilizzata* di numeri reali (cfr. 2.2.1) che individua il numero reale a allora $\{a_n\}$ è convergente e il suo limite è a. Non è vero il viceversa; per esempio, la successione

$$a_n = \begin{cases} 1 + 10^{-n} & \text{per } n \text{ pari} \\ 1 - 10^{-n} & \text{per } n \text{ dispari} \end{cases}$$

non è stabilizzata, ma è convergente a 1.

Il seguente teorema mette in relazione limiti di funzioni e limiti di successioni.

■ **Teorema 3.6** - *Sia data $f : X \to \mathbb{R}$ e sia x_0 un punto di accumulazione per X. Allora si ha:*

$$\lim_{x \to x_0} f(x) = l \tag{3.1}$$

se e solo se, per ogni *successione $\{a_n\}$, a valori in $X \setminus \{x_0\}$ e convergente a x_0, risulta*

$$\lim_{n \to +\infty} f(a_n) = l\ . \tag{3.2}$$

Dimostrazione - Siano x_0 ed $l \in \mathbb{R}$. Mostriamo che da (3.1) segue (3.2). Infatti, se vale (3.1), fissato $\varepsilon > 0$, esiste un $\delta > 0$ tale che:

$$|f(x) - l| < \varepsilon \text{ se } x \in X \text{ e } 0 < |x - x_0| < \delta\ .$$

Ma , se $\{a_n\}$ è convergente a x_0, in corrispondenza di questo δ esiste un N tale che: se $n > N$ risulta $0 < |a_n - x_0| < \delta$ e pertanto $|f(a_n) - l| < \varepsilon$. Vale perciò la (3.2).

Viceversa, mostriamo che da (3.2) segue (3.1). Per assurdo: valga (3.2) ma (3.1) sia falsa; allora esiste un $\varepsilon > 0$ tale che, per ogni $\delta > 0$, si può trovare un punto $a \in X$ con $0 < |a - x_0| < \delta$ e $|f(a) - l| \geq \varepsilon$. Diamo a δ successivamente i valori $1, 1/2, 1/3, \cdots, 1/n \cdots$ e siano a_n i corrispondenti punti di X tali che $0 < |a_n - x_0| < \dfrac{1}{n}$ e $|f(a_n) - l| \geq \varepsilon$; la successione $\{a_n\}$ converge a x_0, ma $f(x_n)$ non tende a l contro l'ipotesi (3.2). La dimostrazione si estende senza difficoltà ai casi in cui x_0 o l non siano finiti. □

Questo teorema ci permette di ricavare molte proprietà dei limiti per le funzioni da analoghe proprietà dei limiti per le successioni. Qualche applicazione sarà data nel prossimo paragrafo.

È anche utile nel caso si voglia dimostrare che una funzione f non ha limite per $x \to x_0$: sarà sufficiente a tale scopo esibire due successioni $\{x_n\}$ e $\{x'_n\}$ entrambe tendenti ad x_0 e tali che $f(x_n) \to l_1$, $f(x'_n) \to l_2$ con $l_1 \neq l_2$. Un esempio è dato negli esercizi.

3.2 Confronti

Successioni infinitesime (cioè convergenti a zero) o infinite (cioè divergenti) possono essere confrontate tra loro come si è fatto per le funzioni; valgono definizioni analoghe a quelle date in 2.5.

Esempi

3.7. Confrontiamo le successioni $\{\sqrt{n}\}$ e $\{a^n\}$ $(a > 1)$. Posto $a = 1 + h$ $(h > 0)$ abbiamo:

$$\frac{\sqrt{n}}{a^n} = \frac{\sqrt{n}}{(1+h)^n} \leq \frac{\sqrt{n}}{1+nh} \leq \frac{1}{h\sqrt{n}} \to 0 .$$

Dunque $\{a^n\}$ è un infinito di ordine superiore a $\sqrt{n}$.

3.8. Verifichiamo che $\lim\limits_{n\to+\infty} na^{-n} = 0$ $(a > 1)$. Ci si riduce al caso precedente scrivendo $na^{-n} = (\sqrt{n}\, b^{-n})^2$ con $b = \sqrt{a}$.

Più in generale, osservando che si può scrivere $n^k a^{-n} = (nb^{-n})^k$ con $b = a^{1/k}$, risulta $\lim\limits_{n\to+\infty} n^k a^{-n} = 0$ per ogni k intero ≥ 1.

E infine, osservando che, per ogni $\alpha > 0$, $n^{[\alpha]} \leq n^\alpha \leq n^{[\alpha]+1}$ possiamo concludere (applicando i risultati precedenti e il teorema del confronto)

$$\lim_{n\to+\infty} n^\alpha a^{-n} = 0 . \tag{3.3}$$

Si noti che la (3.3) vale banalmente anche per $\alpha \leq 0$.

3.9. Se, al posto di n, c'è una successione qualsiasi $\{p_n\}$ divergente a $+\infty$, i risultati precedenti sussistono ancora:

$$\lim_{n\to+\infty} \frac{p_n^\alpha}{a^{p_n}} = 0 \ , \ \ \alpha \in \mathbb{R}, \quad a > 1 . \tag{3.4}$$

Si comincia col mostrare che $\sqrt{p_n}\,a^{-p_n} \to 0$; infatti, posto $a = 1 + h$, si ha:

$$\frac{\sqrt{p_n}}{a^{p_n}} = \frac{\sqrt{p_n}}{(1+h)^{p_n}} \leq \frac{\sqrt{p_n}}{(1+h)^{[p_n]}} \leq \frac{\sqrt{p_n}}{(1+[p_n]h)} \leq \frac{1}{h\sqrt{[p_n]}} + \frac{1}{h[p_n]} \to 0 .$$

Si procede poi come in 3.8.

3.10. Poniamo, nella (3.4), $p_n = \log_b n$ (con $b > 1$); allora, tenendo conto che $\log_b n = \log_a n \cdot \log_b a$, posto $\beta = \log_b a$, abbiamo

$$\lim_{n\to+\infty} \frac{(\log_b n)^\alpha}{n^\beta} = 0 \tag{3.5}$$

relazione valida per ogni $\alpha \in \mathbb{R}$, $\beta > 0$, $b > 1$.

I limiti sopra illustrati descrivono la "velocità" con cui gli esponenziali (con base > 1), le potenze, i logaritmi (con base > 1) vanno all'∞: questi vanno più lentamente di qualsiasi potenza, le potenze più lentamente di qualsiasi esponenziale.

Infiniti di ordine superiore agli esponenziali sono illustrati dai seguenti due limiti che proponiamo per esercizio:

$$\lim_{n\to+\infty} \frac{a^n}{n!} = 0 \quad (a > 1) \tag{3.6}$$

$$\lim_{n\to+\infty} \frac{n!}{n^n} = 0 \ . \tag{3.7}$$

[Suggerimento: per la (3.6), posto $h = a/([a]+1)$, risulta $0 < h < 1$ e, scrivendo $n = [a] + m$, abbiamo

$$\frac{a^n}{n!} = \frac{a}{1}\cdot\frac{a}{2}\cdots\frac{a}{[a]}\cdot\frac{a}{[a]+1}\cdots\frac{a}{[a]+m} < a^{[a]}h^m \ ;$$

per la (3.7) si osservi che

$$\frac{n!}{n^n} = \frac{1}{n}\cdot\frac{2}{n}\cdot\frac{3}{n}\cdots\frac{[\sqrt{n}]}{n}\cdot\frac{([\sqrt{n}]+1)}{n}\cdots\frac{n}{n} \leq \left(\frac{1}{\sqrt{n}}\right)^{[\sqrt{n}]} \ ] \ .$$

Utilizzando il teorema 3.6 i confronti fra esponenziali, potenze, logaritmi si estendono dalle successioni alle funzioni. Si ottiene così, dalla (3.4):

$$\lim_{x\to+\infty} \frac{x^\alpha}{a^x} = 0 \qquad \alpha \in \mathbb{R}, \ a > 1 \ . \tag{3.8}$$

Se nella (3.8) poniamo $x = \log_b t$ e utilizziamo il teorema 2.9 otteniamo (scrivendo β in luogo di $\log_b a$)

$$\lim_{x\to+\infty} \frac{(\log_b x)^\alpha}{x^\beta} = 0 \qquad \alpha \in \mathbb{R}, \beta > 0, b > 1 \ . \tag{3.9}$$

Osservazione 3.2 - In 2.3.2 e 2.3.3 abbiamo introdotto le operazioni di elevamento a potenza con esponente reale e di logaritmo utilizzando *successioni stabilizzate* di numeri razionali; le stesse operazioni si possono definire anche utilizzando *successioni convergenti* di numeri razionali (ma il concetto di successione stabilizzata è più elementare di quello di successione convergente!). Accenniamo al procedimento per definire a^α (con $a > 1$ e $\alpha \in \mathbb{R}$).

Sia $\{p_n\}$ una successione crescente di numeri razionali convergente ad α (si scrive : $p_n \uparrow \alpha$). La successione $\{a^{p_n}\}$ è anch'essa crescente, limitata e perciò convergente; sia l il suo limite. Per poter definire $a^\alpha := l$ occorre verificare che, se scegliamo un'altra successione $\{q_n\}$, $q_n \uparrow \alpha$, si ha ancora $a^{q_n} \uparrow l$.

Si ha:

$$a^{p_n} - a^{q_n} = a^{q_n}(a^{p_n - q_n} - 1) \ .$$

Poiché $a^{q_n} \to l$ e $a^{p_n - q_n} \to 1$ (essendo $p_n - q_n \to \alpha - \alpha = 0$), risulta che $a^{p_n} - a^{q_n} \to 0$; dunque a^{p_n} e a^{q_n} tendono allo stesso limite.

Possiamo quindi definire $a^\alpha := \lim_{n \to +\infty} a^{p_n}$ dove p_n è una *qualunque* successione di razionali tale che $p_n \uparrow \alpha$.

Si possono poi verificare senza difficoltà le proprietà E_0-E_5.

Esercizi

1. Usare la definizione di limite per verificare che

a) $\dfrac{n^2}{n^2+1} \to 1$

b) $\sqrt{n-1} \to +\infty$

c) $\sqrt[n]{n^2} \to 1$

d) $\dfrac{n^{10}}{10^n} \to 0$.

2. Sia $\{a_n\}$ una successione a valori positivi.

a) Dimostrare che, se esiste un numero k, $0 < k < 1$, tale che, definitivamente, $\sqrt[n]{a_n} < k$ allora $a_n \to 0$.

b) Dimostrare che, se esiste un numero $k, k > 1$, tale che, definitivamente, $\sqrt[n]{a_n} > k$ allora $a_n \to +\infty$.

3. Sia $\{a_n\}$ una successione a valori positivi tale che $\sqrt[n]{a_n} \to l \in \mathbf{R}^*$.
Utilizzando il risultato dell'esercizio 2 dimostrare che

a) se $l < 1$ allora $a_n \to 0$

b) se $l > 1$ allora $a_n \to +\infty$.

Esaminare il caso $l = 1$.

4.* Sia $\{a_n\}$ una successione a valori positivi.

a) Dimostrare che se esiste un numero $k, 0 < k < 1$, tale che, definitivamente, $\dfrac{a_{n+1}}{a_n} < k$ allora $a_n \to 0$.

b) Dimostrare che se esiste un numero $k, k > 1$, tale che, definitivamente, $\dfrac{a_{n+1}}{a_n} > k$ allora $a_n \to +\infty$.

5. Sia $\{a_n\}$ una successione a valori positivi tale che $\dfrac{a_{n+1}}{a_n} \to l \in \mathbf{R}^*$.
Utilizzando il risultato dell'esercizio 4 dimostrare che

a) se $l < 1$ allora $a_n \to 0$

b) se $l > 1$ allora $a_n \to +\infty$.

Esaminare il caso $l = 1$.

6. Utilizzare l'esercizio 5 per ritrovare che $\frac{a^n}{n!} \to 0$ ($\forall a \in \mathbb{R}$).

7. Calcolare i limiti delle seguenti successioni

a) $\frac{\sqrt[3]{n} + \log_2 n + 1}{n^2 + 1}$

b) $\sqrt{n} \cdot \sin \frac{1}{n^\alpha}$ ($\alpha \in \mathbb{R}_+$)

c) $\frac{n!}{(n+1)! - (n-1)!}$

d) $\sqrt[3]{n^6 - n^\alpha + 1} - n^2$ ($\alpha \in \mathbb{R}_+$).

8. Calcolare i limiti delle seguenti successioni:

a) $\frac{n^\alpha}{1 + 2 + \cdots + n}$ ($\alpha \in \mathbb{R}_+$)

b) $\frac{1 + 3 + 5 + \cdots + 2n + 1}{1 + 2 + 3 + \cdots + n}$.

9.* Dimostrare, utilizzando il teorema di confronto 3.3, che

a) $\lim_{n \to +\infty} \left(\frac{1}{n^2+1} + \cdots + \frac{1}{n^2+n} \right) = 0$

b) $\lim_{n \to +\infty} \left(\frac{1}{\sqrt{n^2+1}} + \cdots + \frac{1}{\sqrt{n+n^2}} \right) = 1$.

10. Dimostrare che $\lim_{x \to 0_+} \cos \frac{1}{x}$ non esiste utilizzando il teorema 3.6.

[esibire 2 successioni $\{x_n\}, \{x'_n\}$ tali che $x_n \to 0_+, x'_n \to 0_+$ e $\cos \frac{1}{x_n} \to l, \cos \frac{1}{x'_n} \to l'$ con $l \neq l'$].

11. Calcolare i limiti delle seguenti successioni

a) $\frac{\log_a n + 1}{\sqrt{n} + 1}$

b) $\frac{n + 2^n}{n + n!}$

c*) $\frac{(2n)!}{n^n}$.

12. Dimostrare che $\forall \alpha > 0$,

$$\lim_{x \to 0_+} x^\alpha \log_2 x = 0 \ .$$

13.** Dimostrare che se $a_n \to l \in \mathbb{R}$ allora

$$\frac{a_1 + a_2 + \cdots + a_n}{n} \to l \ .$$

È vero il viceversa? [Considerare $a_n = (-1)^n$].

Utilizzando ora il risultato precedente dimostrare che se $a_n \to l \in \mathbb{R}$, $a_n > 0\ \forall n > 1$, allora $\sqrt[n]{a_1 \cdot a_2 \cdots a_n} \to l$.

È vero il viceversa? [Considerare $a_n = 2 + (-1)^n$].

14. Sia data la seguente successione per ricorrenza

$$s_0 = 1 \qquad s_{n+1} = \sqrt{s_n + 2}\ .$$

Dimostrare che:

i) s_n è monotona crescente

ii) s_n è limitata: $0 < s_n < 2$

iii) dedurre che $\lim\limits_{n\to+\infty} s_n$ esiste ed è 2.

15. Dimostrare che la successione di Fibonacci F_n definita da $F_0 = 0$, $F_1 = 1$, $F_{n+1} = F_n + F_{n-1}$ è divergente a $+\infty$.

3.3 Il numero "e"; alcuni limiti notevoli

Dimostreremo, in questo paragrafo, il seguente

■ **Teorema 3.7** - *La successione*

$$a_n = \left(1 + \frac{1}{n}\right)^n \qquad n = 1, 2, \ldots \tag{3.10}$$

è convergente.

Osservazione 3.3 - Il limite della successione (3.10) viene universalmente indicato con la lettera "*e*". Esso, come vedremo, è un numero *irrazionale* la cui rappresentazione decimale inizia così:

2. 7 1 8 2 8 1 8 2 8 4 ⋯

Questo numero, che è secondo in quanto a popolarità solo a π, entra in moltissime questioni, non solo di matematica. Tanto per cominciare, esso viene spessissimo utilizzato (per ragioni che vedremo), come base dei logaritmi, i quali, quando si usa questa base, vengono detti *naturali* o *neperiani* (*) e indicati semplicemente col simbolo log (oppure ln) senza indicazione della base (quando si usa la base 10 si scrive talvolta Log invece di $\log_{10}$).

Il numero "*e*" è popolare anche tra gli economisti per il seguente motivo.

Supponiamo di possedere un capitale (per esempio, 1 milione di dollari) e di investirlo al tasso t di interesse annuale (cioè con una rendita di t milioni all'anno).

(*) Dal nome del matematico John Napier (1550–1617), inventore dei logaritmi.

Se l'interesse viene pagato annualmente, dopo 1 anno il capitale posseduto sarà $1 + t$. Se l'interesse viene calcolato mensilmente avremo:

dopo il primo mese un capitale pari a $1 + \frac{t}{12}$;

dopo il secondo mese pari a $1 + \frac{t}{12} + \frac{t}{12}\left(1 + \frac{t}{12}\right) = \left(1 + \frac{t}{12}\right)^2$;

dopo il terzo mese pari a $\left(1 + \frac{t}{12}\right)^2 \frac{t}{12} + \left(1 + \frac{t}{12}\right)^2 = \left(1 + \frac{t}{12}\right)^3$;

alla fine dell'anno avremo un capitale pari a $\left(1 + \frac{t}{12}\right)^{12}$.

Se l'interesse viene calcolato ogni n–esimo di anno, avremo alla fine un capitale pari a

$$\left(1 + \frac{t}{n}\right)^n .$$

Se $t = 1$ (rendita del 100%!) otteniamo esattamente la successione che definisce "e".

Far tendere n ad ∞ significa calcolare l'interesse dopo frazioni di anno sempre più piccole fino ad arrivare a calcolarlo con continuità. Il capitale che si ottiene alla fine dell'anno in quest'ultimo caso è esattamente pari ad "e".

Dimostrazione del Teorema 3.7 - Si svolge in due passi:

i) mostriamo che $\{a_n\}$ è monotona crescente

ii) mostriamo che $\{a_n\}$ è limitata.

Di conseguenza (cfr. proposizione 3.5) $\{a_n\}$ converge al suo estremo superiore.

i) Risulta $a_1 = 2$ e, per $n \geq 2$,

$$\frac{a_n}{a_{n-1}} = \frac{(1 + 1/n)^n}{(1 + 1/(n-1))^{n-1}} = \left(\frac{n+1}{n}\right)^n \left(\frac{n-1}{n}\right)^n \Big/ \frac{n-1}{n} = \frac{(1 - 1/n^2)^n}{1 - 1/n} .$$

Per la disuguaglianza di Bernoulli $(1 - 1/n^2)^n > 1 - 1/n$ e perciò $a_n > a_{n-1}$; risulta allora, $a_n \geq 2 \; \forall n = 1, 2, \cdots$

ii) Consideriamo la successione $\{b_n\}$ così definita:

$$b_n = \left(1 + \frac{1}{n}\right)^{n+1} ;$$

risulta $b_1 = 4$ e per $n \geq 2$, (ragionando come in *i*))

$$\frac{b_n}{b_{n-1}} = \frac{(1 + 1/n)^{n+1}}{(1 + 1/(n-1))^n} = \frac{1 + 1/n}{(1 + 1/(n^2 - 1))^n} .$$

Ma

$$\left(1+\frac{1}{n^2-1}\right)^n > 1+\frac{n}{n^2-1} > 1+\frac{n}{n^2} = 1+\frac{1}{n}$$

cosicché $b_n < b_{n-1}$; dunque $b_n \leq 4 \ \forall n = 1, 2, \cdots$.

D'altra parte $b_n > a_n$, da cui si ricava

$$2 \leq a_n < 4 \ . \ \square$$

Osservazione 3.4 - Dalla dimostrazione precedente risulta che $\{a_n\}$ *converge ad e per difetto.* Risulta anche che $\{b_n\}$ *converge ad e per eccesso*; infatti $b_n = a_n(1+1/n)$ e perciò $\lim_{n\to+\infty} b_n = e$; ma essendo $\{b_n\}$ decrescente, essa converge al suo estremo inferiore. Prendendo, per esempio, $n = 5$, si trova:

$$2.4883 < e < 2.9860 \ .$$

Infine lo scarto $b_n - a_n$ può essere così stimato:

$$b_n - a_n = b_n \frac{1}{n+1} \leq \frac{4}{n+1} \ .$$

Questa disuguaglianza indica che ... la definizione di e non è adeguata al calcolo delle sue cifre decimali. Infatti, arrestandosi al termine n–esimo della successione $\{a_n\}$ ($\{b_n\}$) otteniamo una approssimazione per difetto (per eccesso) di e con un errore inferiore a $4/(n+1)$. Tale procedimento di calcolo approssimato del numero e è perciò troppo lento, poiché, ad esempio, prendendo $n = 400$, abbiamo ancora incertezza sulla seconda cifra decimale! Per il calcolo di e si usano infatti metodi diversi che vedremo in seguito.

Corollario 3.8 - *Se* $\{x_n\}$ *è una successione tale che, per* $n \to +\infty$,

i) $x_n \to +\infty$ *oppure ii)* $x_n \to -\infty$ *oppure iii)* $x_n \to \infty$, *sempre si deduce*

$$\lim_{n\to+\infty}\left(1+\frac{1}{x_n}\right)^{x_n} = e \ . \tag{3.11}$$

Dimostrazione -

i) Poniamo

$$\alpha_n = \left(1+\frac{1}{[x_n]+1}\right)^{[x_n]} \quad \text{e} \quad \beta_n = \left(1+\frac{1}{[x_n]}\right)^{[x_n]+1}$$

Poiché $[x_n] \leq x_n \leq [x_n]+1$, si ha

$$\alpha_n \leq \left(1+\frac{1}{x_n}\right)^{x_n} \leq \beta_n \ .$$

Ma

$$\alpha_n = \left(1+\frac{1}{[x_n]+1}\right)^{[x_n]+1}\left(1+\frac{1}{[x_n]+1}\right)^{-1} .$$

Per cui, applicando il teorema 3.7, $\alpha_n \to e$; analogamente $\beta_n \to e$; la (3.11) segue allora dal teorema del confronto.

ii) $x_n \to -\infty$; posto $y_n = -x_n$ risulta $y_n \to +\infty$ e

$$\left(1+\frac{1}{x_n}\right)^{x_n} = \left(1-\frac{1}{y_n}\right)^{-y_n} = \left(\frac{y_n}{y_n-1}\right)^{y_n} = \left(1+\frac{1}{y_n-1}\right)^{y_n-1}\left(1+\frac{1}{y_n-1}\right)$$

e la (3.11) segue per quanto dimostrato in *i)*.

iii) $x_n \to \infty$ (esistono infiniti $x_n > 0$ e infiniti $x_n < 0$). Si consideri la successione degli $x_n > 0$ e quella degli $x_n < 0$; per la prima, divergente a $+\infty$, possiamo applicare il punto *i)*; per la seconda, divergente a $-\infty$, si applica il punto *ii)*; allora per la successione completa, ragionando in base alla definizione di successione convergente, si conclude che vale ancora la (3.11). □

Corollario 3.9 - *Sia x tendente a $+\infty$, oppure $-\infty$, oppure ∞. Allora*

$$\left(1+\frac{1}{x}\right)^{x} \to e . \tag{3.12}$$

L'affermazione segue dal teorema 3.6 e dal precedente corollario.

Dalla (3.12), utilizzando ripetutamente il teorema 2.9, si deducono i seguenti limiti, fondamentali per l'analisi matematica.

Proposizione 3.10 -

$$\left(1+\frac{\theta}{x}\right)^{x} \to e^{\theta} \quad \textit{per } x \to \begin{matrix} +\infty \\ -\infty \\ \infty \end{matrix} \quad \forall\theta \in \mathbb{R} \tag{3.13}$$

$$(1+\theta y)^{1/y} \to e^{\theta} \quad \textit{per } y \to 0 \quad \forall\theta\in\mathbb{R} \tag{3.14}$$

$$\frac{\log(1+y)}{y} \to 1 \quad \textit{per } y \to 0 \tag{3.15}$$

$$\frac{a^y-1}{y} \to \log a \quad \textit{per } y \to 0 \quad \forall a>0 \tag{3.16}$$

$$\frac{(1+y)^{\theta}-1}{y} \to \theta \quad \textit{per } y \to 0 \quad \forall\theta\in\mathbb{R} . \tag{3.17}$$

Dimostrazione -

(3.13) Se $\theta = 0$ è ovvia. Se $\theta \neq 0$ si ponga $t = x/\theta$ e si applichi (3.12).
(3.14) Dalla (3.13) ponendo $x = 1/y$.
(3.15) Dalla (3.14), con $\theta = 1$, passando ai logaritmi.

(3.16) Posto $z = a^y - 1$, risulta $z \to 0$ e $y = \log_a(z+1) = \dfrac{1}{\log a}\log(z+1)$. Perciò $\dfrac{a^y - 1}{y} = \dfrac{z}{\log(z+1)}\log a$ e l'affermazione segue dalla (3.15).

(3.17) Posto $z = (1+y)^\theta - 1$, cioè $\log(1+z) = \theta\log(1+y)$, risulta $z \to 0$ e

$$\frac{(1+y)^\theta - 1}{y} = \frac{z}{\log(1+z)} \cdot \frac{\log(1+y)}{y}\theta$$

e l'affermazione segue dalla (3.15). □

A conclusione di questo paragrafo presentiamo, senza dimostrazione, una formula che risulta di grande utilità in svariate questioni sia teoriche che pratiche.

Proposizione 3.11 - (Formula di De Moivre–Stirling)

$$n! = n^n e^{-n}\sqrt{2\pi n}\; e^{\theta_n/12n} \quad \text{con } 0 < \theta_n < 1 \,. \tag{3.18}$$

La (3.18) ci fornisce un valore approssimato di $n!$ con un errore relativo che tende a zero per $n \to +\infty$.

Possiamo anche scrivere

$$n! \sim n^n e^{-n}\sqrt{2\pi n} \tag{3.19}$$

o anche, passando ai logaritmi,

$$\log n! = n\log n - n + \frac{1}{2}\log n + \log\sqrt{2\pi} + o(1)\,. \tag{3.20}$$

Esercizi

16. Calcolare i seguenti limiti

a) $\lim_{x\to 0_+}(1+x)^{1/x^\alpha}\ (\alpha \in \mathbb{R}_+)$

b) $\lim_{x\to+\infty}(1 + 2\sin\frac{1}{x})^x$

c) $\lim_{x\to 0}\dfrac{\sqrt[3]{1-3x} - 1}{x}$

d) $\lim_{x\to 0}\dfrac{5^x - 7^x}{x}$

e) $\lim_{x\to+\infty}\left(\dfrac{3+x^\alpha}{2+x}\right)^x\ (\alpha \in \mathbb{R})$.

17.** Dimostrare che, se $a_n \to a$, allora

$$\left(1+\frac{a_1}{n}\right)\cdot\left(1+\frac{a_2}{n}\right)\cdots\left(1+\frac{a_n}{n}\right) \to e^a\,.$$

18. Calcolare i limiti delle seguenti successioni

a) $\dfrac{(2n)!}{(n!)^2}$

b) $\dfrac{(2e)^n\sqrt{n^2+1}\cdot n!}{(2n)^n\cdot n^2}$

c) $\sqrt[n]{n!}$

d) $\left(1+\dfrac{1}{n!}\right)^{n^n}$.

3.4 Esistenza del limite. Massimo e minimo limite

Vogliamo indagare ora, con maggiore profondità di quanto abbiamo fatto per le funzioni, circa l'esistenza del limite. Cominciamo col definire un importante concetto: quello di sottosuccessione.

Sia $\{a_n\}$ una successione. Supponiamo di selezionare, mediante una qualche legge, alcuni termini della successione $\{a_n\}$ e di ordinarli in una nuova successione $\{b_n\}$, in modo che si succedano nello stesso ordine che avevano nella $\{a_n\}$; quello che otteniamo è una *sottosuccessione*, o *successione parziale* o *estratta* dalla $\{a_n\}$. Più precisamente, abbiamo la seguente

Definizione 3.1 - *Si chiama sottosuccessione della successione $\{a_n\}$ una successione $\{b_n\}$ tale che:*

$$b_n = a_{k_n} \qquad \forall n \in \mathbb{N}$$

dove $\{k_n\}$ è una successione strettamente crescente di interi naturali.

***Esempio* 3.11** - Sia $a_n = n \sin \dfrac{\pi}{2} n$. Allora sono sottosuccessioni estratte dalla $\{a_n\}$ le successioni

$$\{0,0,0,\cdots,0,\cdots\} \qquad k_n = 2n$$

$$\{1,-3,5,-7,9,\cdots\} \qquad k_n = 2n+1$$

$$\{1,9,25,49,81,\cdots\} \qquad k_n = (2n+1)^2 \; .$$

Il teorema seguente ci dà una *condizione necessaria e sufficiente* perché una successione abbia limite.

■ **Teorema 3.12** - *Sia data $\{a_n\}$ e sia $l \in \mathbb{R}^*$ oppure $l \in \dot{\mathbb{R}}$; $\{a_n\}$ ha limite l se e solo se ogni sottosuccessione di $\{a_n\}$ (definitivamente diversa da $\{a_n\}$) ha limite l.*

Dimostrazione - Supponiamo che ogni sottosuccessione definitivamente diversa da $\{a_n\}$ abbia limite l. Neghiamo la tesi e mostriamo che risulta falsa l'ipotesi. Sia $l \in \mathbb{R}$. Se $\{a_n\}$ non ammette come limite l, esiste un $\varepsilon > 0$ tale che, per ogni n, esiste un $n' \geq n$ tale che

$|a_{n'} - l| \geq \varepsilon$. Preso allora $n = 1$, sia k_1 l'intero n' corrispondente, preso poi $n = k_1 + 1$, sia k_2 l'intero n' corrispondente, e così via. Si costruisce così una successione crescente di interi k_n tali che $|a_{k_n} - l| \geq \varepsilon \ \ \forall n$; allora la sottosuccessione $\{a_{k_n}\}$ così costruita non è convergente a l. Dimostrazione analoga per $l = +\infty, -\infty, \infty$.

Viceversa, sia $l \in \mathbb{R}$. Per ipotesi, fissato $\varepsilon > 0$, esiste N tale che: $|a_n - l| < \varepsilon$ per $n \geq N$. Sia $\{b_n\}$ una sottosuccessione di $\{a_n\}$ con $b_n = a_{k_n}$. Poiché k_n è crescente, sarà $k_n \geq n$ e perciò, a maggior ragione, $|a_{k_n} - l| < \varepsilon$ per $n \geq N$, cioè l'asserto. Analoga è la dimostrazione per $l = +\infty, -\infty, \infty$. □

Definizione 3.2 - *Sia $\{a_n\}$ una successione. Un elemento $l \in \mathbb{R}^*$ si dice* **valore limite** *di $\{a_n\}$ se si verifica una delle seguenti circostanze tra loro equivalenti:*

i) esiste una sottosuccessione di $\{a_n\}$ che ammette l come limite

ii) l è punto di accumulazione per l'insieme $\{a_n\}_{n \in \mathbb{N}}$, oppure esistono infiniti indici n_k per cui $a_{n_k} = l$.

L'insieme dei valori limite di $\{a_n\}$ si dice **classe limite** di $\{a_n\}$. Sottolineiamo che la classe limite di una successione è un sottoinsieme di $\mathbb{R}^*$.

Prima di mostrare l'equivalenza delle due definizioni date, consideriamo i seguenti esempi.

***Esempio* 3.12** - La classe limite della successione $\{(-1)^n\}$ è $\{-1, +1\}$.

La classe limite della successione $\{(-2)^n\}$ è $\{-\infty, +\infty\}$.

La classe limite della successione $\{1/n\}$ è $\{0\}$.

La classe limite della successione dei numeri razionali coincide con $\mathbb{R}^*$.

Mostriamo l'equivalenza di *i*) e *ii*). L'implicazione *i*)$\Rightarrow$ *ii*) è evidente. Viceversa, se l è punto di accumulazione per l'insieme $\{a_n\}_{n \in \mathbb{N}}$, fissato $\varepsilon_1 > 0$ si prenda il primo elemento della successione, diciamolo a_{k_1}, appartenente all'intervallo $(l - \varepsilon_1, l + \varepsilon_1)$; fissato poi $\varepsilon_2 = l - a_{k_1}$, si prenda il primo elemento, diciamolo a_{k_2}, appartenente a $(l - \varepsilon_2, l + \varepsilon_2)$; così continuando si costruisce una successione crescente di interi $\{k_n\}$, tale che $a_{k_n} \to l$ (vedi più avanti, per una generalizzazione, la proposizione 4.2); se infine l non è punto di accumulazione, una sottosuccessione tendente a l è a_{n_k}.

Una notevole proprietà della classe limite di una successione è espressa dal seguente

▪ **Teorema 3.13** - *La classe limite di una successione è chiusa in $\mathbb{R}^*$, cioè ogni punto di accumulazione della classe limite è anch'esso un valore limite.*

Dimostrazione - Sia λ un punto di accumulazione di valori limite; $\forall \varepsilon > 0$, nell'intervallo $(\lambda - \varepsilon, \lambda + \varepsilon)$ cadono infiniti valori limite; sia l_1 uno di questi: $\lambda - \varepsilon < l_1 < \lambda + \varepsilon$ e prendiamo un intorno di l_1 tutto contenuto in $(\lambda - \varepsilon, \lambda + \varepsilon)$; in esso sono contenuti infiniti elementi della successione e perciò in $(\lambda - \varepsilon, \lambda + \varepsilon)$ sono contenuti infiniti elementi della successione; per l'arbitrarietà di ε, anche λ è un valore limite. □

È evidente che, se una successione ammette come limite $l \in \mathbb{R}^*$, la sua classe limite contiene un unico elemento, l appunto.

Può la classe limite di una successione $\{a_n\}$ essere vuota? La risposta è negativa; infatti, se l'immagine della successione è finita, cioè se ci sono soltanto un numero finito di valori a_n distinti, almeno uno di essi dovrà comparire infinite volte nella successione, e pertanto è un valore limite; se poi l'insieme $\{a_n\}$ è infinito, il teorema di Bolzano–Weierstrass ci assicura che esiste in $\mathbb{R}^*$ almeno un punto di accumulazione di questo insieme, cioè un valore limite della successione.

Dunque la classe limite $\Lambda \in \mathbb{R}^*$ di una successione $\{a_n\}$ è un insieme *chiuso* (in $\mathbb{R}^*$) e *non vuoto*; perciò esisterà in esso un elemento *massimo* $\bar{l}$ e un elemento *minimo* $\underline{l}$ (eventualmente coincidenti).

Si osservi che, se $\{a_n\}$ non è limitata superiormente, tale elemento massimo è $+\infty$, se non è limitata inferiormente, l'elemento minimo è $-\infty$.

Si noti che parliamo di massimo e di minimo, e non di estremo superiore e inferiore, poiché questi elementi appartengono a Λ, essendo Λ chiuso in $\mathbb{R}^*$.

A questo punto sono naturali le seguenti definizioni.

Definizione 3.3 - *Si dice* **limite superiore** *o* **massimo limite** *di una successione il massimo dei valori limite:* esso è $+\infty$ quando la successione non è limitata superiormente, è $-\infty$ quando la successione diverge a $-\infty$, è un numero reale in ogni altro caso.

Analogamente *si dice* **limite inferiore** *o* **minimo limite** *di una successione il minimo dei valori limite:* esso è $-\infty$ quando la successione non è limitata inferiormente, è $+\infty$ se la successione diverge a $+\infty$, è un numero reale in ogni altro caso.

Le notazioni usate per indicare il massimo limite sono:

$$\overline{\lim_{n\to+\infty}} a_n \qquad \limsup_{n\to+\infty} a_n \qquad \max\lim_{n\to+\infty} a_n$$

e analogamente per il minimo limite

$$\underline{\lim_{n\to+\infty}} a_n \qquad \liminf_{n\to+\infty} a_n \qquad \min\lim_{n\to+\infty} a_n \ .$$

Dalle definizioni date immediatamente si ricavano le seguenti proprietà:

i) $\underline{\lim} a_n \leq \overline{\lim} a_n$

ii) $\{a_n\}$ ha limite $\Leftrightarrow$ $\underline{\lim} a_n = \overline{\lim} a_n$

iii) da ogni successione limitata superiormente (inferiormente) si può estrarre una sottosuccessione convergente al massimo limite (minimo limite).

Esempi

3.13. La successione $\{\sin \frac{n\pi}{4}\}$ è irregolare e la sua classe limite è $\{0, \sqrt{2}/2, 1, -1, -\sqrt{2}/2\}$. Perciò $\underline{\lim}_{n\to+\infty} \sin\frac{n\pi}{4} = -1$, $\overline{\lim}_{n\to+\infty} \sin\frac{n\pi}{4} = +1$.

3.14. La successione $\{\operatorname{tg}\frac{n\pi}{3}\}$ ha come classe limite $\{-\sqrt{3}, 0, \sqrt{3}\}$

3.15. La successione $\{a_n\}$ dove $a_n = \sqrt{n+1} - [\sqrt{n}]$ è irregolare. 0 è un valore limite di $\{a_n\}$: infatti la sottosuccessione $\{a_{k_n}\}$ con $k_n = n^2$ tende a 0 ($\lim_{n\to+\infty} \sqrt{n^2+1} - n = 0$). Anche 1 è un valore limite di $\{a_n\}$ perché la sottosuccessione $\{a_{h_n}\}$ con $h_n = n^2 - 1$ tende a 1 (di fatto è la successione costante $\{1\}$); inoltre risulta $0 < a_n \leq 1$ per ogni n. Ciò consente di concludere che $\underline{\lim}_{n\to+\infty} a_n = 0$, $\overline{\lim}_{n\to+\infty} a_n = 1$.

In questo paragrafo ci eravamo posti il problema di indagare circa l'esistenza del limite di una successione a valori in $\mathbb{R}$. Sappiamo bene che non tutte le successioni ammettono limite, anzi, possiamo dire che quelle che ammettono limite sono un po' un'eccezione nell'insieme di tutte le successioni. Siamo però arrivati a trovare due elementi di $\mathbb{R}^*$ che sono, in qualche misura, sostitutivi del concetto di limite: il massimo e il minimo limite, i quali hanno però la buona proprietà di esistere per *ogni* successione.

Questi due elementi possono essere caratterizzati anche da una proprietà che, in un certo senso, ricorda quella dell'estremo superiore e inferiore.

Proposizione 3.14 - *Se $\bar{l} \in \mathbb{R}$ è il $\overline{\lim}$ di $\{a_n\}$, allora $\bar{l} = \inf K$, dove K è la classe dei numeri definitivamente maggioranti di $\{a_n\}$.*

Dimostrazione - Se $\bar{l} = \inf K$, nell'intervallo $(\bar{l} - \varepsilon, \bar{l} + \varepsilon)$, $\forall \varepsilon > 0$, cadono infiniti a_n; perciò $\bar{l}$ è un valore limite; non può esserci un altro valore limite più grande, diciamo $\bar{\bar{l}}$, perché, scegliendo $\delta > 0$ tale che $\bar{\bar{l}} - \delta > \bar{l}$, nell'intervallo $(\bar{\bar{l}} - \delta, \bar{\bar{l}} + \delta)$ cadrebbero infiniti a_n in contrasto con la definizione di $\bar{l}$. □

Analoga proposizione vale per il minimo limite: *se $\underline{l} \in \mathbb{R}$ è il $\underline{\lim}$ di $\{a_n\}$, allora $\underline{l} = \sup H$, dove H è la classe dei numeri definitivamente minoranti di $\{a_n\}$.*

* *Osservazione 3.7* - I concetti di valore limite, classe limite, massimo e minimo limite, sopra illustrati per le successioni, possono essere introdotti anche per le funzioni; ma essi hanno minore importanza nel quadro generale di questo corso, per cui ci limitiamo ad un rapido cenno.

Sia $f : X \to \mathbb{R}$ e $x_0 \in \mathbb{R}^*$ o $\dot{\mathbb{R}}$ un punto di accumulazione di X. Un elemento $l \in \mathbb{R}^*$ si dice *valore limite* di $f(x)$ per $x \to x_0$ se, per ogni coppia di intorni $U(x_0)$ e $V(l)$, esiste almeno un punto $x \in X \cap U(x_0)$ e distinto da x_0 tale che la sua immagine appartiene a $V(l)$.

Equivalentemente, l è valore limite se esiste $\{x_n\} \subset X \setminus \{x_0\}$ tale che $x_n \to x_0$ e $f(x_n) \to l$ (cfr. esercizio 24).

Si dice *classe limite* della funzione $f(x)$ per $x \to x_0$ l'insieme Λ dei valori limite di $f(x)$ per $x \to x_0$.

Si dimostra che la classe limite di $f(x)$ per $x \to x_0$ è un insieme di $\mathbb{R}^*$ chiuso (in $\mathbb{R}^*$) e non vuoto; il suo elemento massimo si dice *massimo limite* (o *limite superiore*) di $f(x)$ per $x \to x_0$; analoga definizione per il *minimo limite* (o *limite inferiore*). Per questi limiti si usano le notazioni:

$$\varliminf_{x\to x_0} f(x)\ ,\ \liminf_{x\to x_0} f(x)\ ,\ \min\lim_{x\to x_0} g(x) \qquad \text{per il minimo limite}$$

$$\varlimsup_{x\to x_0} f(x)\ ,\ \limsup_{x\to x_0} f(x)\ ,\ \max\lim_{x\to x_0} f(x) \qquad \text{per il massimo limite.}$$

Esempi

3.16. $\varliminf_{x\to\infty} \sin x = -1$, $\varlimsup_{x\to\infty} \sin x = 1$; la classe limite di $\sin x$ per $x \to \infty$ è l'intervallo $[-1, +1]$.

3.17. $\varliminf_{x\to 0} \dfrac{1}{x} = -\infty$, $\varlimsup_{x\to 0} \dfrac{1}{x} = +\infty$; la classe limite di $\dfrac{1}{x}$ per $x \to 0$ è $\{-\infty, +\infty\}$.

3.18. $\varliminf_{x\to+\infty} \operatorname{tg} x = -\infty$, $\varlimsup_{x\to+\infty} \operatorname{tg} x = +\infty$; la classe limite di $\operatorname{tg} x$ per $x \to +\infty$ è tutto $\mathbb{R}^*$.

3.19. Sia $f(x) = 0$ per x razionale, $= 1$ per x irrazionale, $x \in \mathbb{R}$ (funzione di Dirichlet). Allora $\forall x_0 \in \mathbb{R}^*$ risulta: $\varliminf_{x\to x_0} f(x) = 0$, $\varlimsup_{x\to 0} f(x) = 1$ e la classe limite di $f(x)$ per $x \to x_0$ è $\{0, 1\}$.

3.20. $\varlimsup_{x\to 0_+} \sin \dfrac{1}{x} = 1$, $\varliminf_{x\to 0_+} \sin \dfrac{1}{x} = -1$; la classe limite di $\sin \dfrac{1}{x}$ per $x \to 0_+$ è l'intervallo $[-1, 1]$.

Esercizi

19. Determinare la classe limite in $\mathbb{R}^*$ delle successioni seguenti:

a) $1 - \cos n\pi$

b) $(-1)^n \dfrac{n+1}{n-1}$

c) $n \sin(n\dfrac{\pi}{2})$

d)* $\sqrt{n} - [\sqrt{n}]$

20. Dimostrare che se $\{x_n\}$ e $\{y_n\}$ sono successioni, si ha:

a) $\underline{\lim}\, x_n + \underline{\lim}\, y_n \le \underline{\lim}\,(x_n + y_n) \le \overline{\lim}\,(x_n + y_n) \le \overline{\lim}\, x_n + \overline{\lim}\, y_n$.

b) $x_n \le y_n$ definitivamente $\Rightarrow \underline{\lim}\, x_n \le \underline{\lim}\, y_n$ e $\overline{\lim}\, x_n \le \overline{\lim}\, y_n$.

21.** Sia $\{a_n\}$ una successione positiva.
Allora

$$\underline{\lim}\, \frac{a_{n+1}}{a_n} \le \underline{\lim}\, \sqrt[n]{a_n} \le \overline{\lim}\, \sqrt[n]{a_n} \le \overline{\lim}\, \frac{a_{n+1}}{a_n}.$$

Dedurre che se $a_{n+1}/a_n \to l$ allora $\sqrt[n]{a_n} \to l$.

[*Cenno di dimostrazione:* la disuguaglianza centrale è ovvia; la prima e la terza hanno dimostrazioni analoghe. Per la prima, poniamo $\underline{\lim}\, a_{n+1}/a_n = \underline{l}$. Si ha $\underline{l} \ge 0$. Se $\underline{l} = 0$ non c'è nulla da dimostrare, sia dunque $\underline{l} > 0$. Per la proposizione 3.14, $\underline{l} = \sup H$, dove H è la classe di numeri definitivamente minoranti di a_{n+1}/a_n. Se $0 < \alpha < \underline{l}$, per $k \ge N$ si avrà $a_{k+1}/a_k > \alpha$, cioè $a_{k+1} > \alpha a_k$. Iterando l'ultima disequazione da $k = N$ sino a $k = n - 1$ si ottiene

$$a_n > \alpha a_{n-1} > \cdots > \alpha^{n-N} a_N$$

da cui $\sqrt[n]{a_n} > \alpha \sqrt[n]{\alpha^{-N} a_N}$ e quindi $\underline{\lim}\, \sqrt[n]{a_n} > \alpha$. Essendo α arbitrario in $(0, \underline{l})$ segue $\underline{\lim}\, \sqrt[n]{a_n} \ge \underline{l}$]

22. Sia $a_n = 1$ per n dispari e $a_n = 2^n$ per n pari. Verificare che

$$\underline{\lim}\, \frac{a_{n+1}}{a_n} = 0\ ,\ \overline{\lim}\, \frac{a_{n+1}}{a_n} = +\infty$$

mentre

$$\underline{\lim}\, \sqrt[n]{a_n} = 1 \text{ e } \overline{\lim}\, \sqrt[n]{a_n} = 2\ .$$

Confrontare con il risultato dell'esercizio 21.

23. Sia $a_n = 2^{n+1}$ per n pari, $a_n = 2^n$ per n dispari.
Verificare che $\sqrt[n]{a_n} \to 2$ mentre $\underline{\lim}\, \dfrac{a_{n+1}}{a_n} = \dfrac{1}{2} \ne 2 = \overline{\lim}\, \dfrac{a_{n+1}}{a_n}$.
Confrontare con il risultato dell'esercizio 21.

24*. Dimostrare che l è un valore limite di f per $x \to x_0 \in \mathbb{R}^*$ se e solo se esiste una successione $\{x_n\} \subset \mathrm{dom}(f) \setminus \{x_0\}$ tale che $x_n \to x_0$ e $f(x_n) \to l$.

3.5 Esistenza del limite finito. Criterio di Cauchy

Il criterio di Cauchy è una condizione necessaria e sufficiente perché una successione ammetta limite in $\mathbb{R}$ (cioè sia *convergente*). Il vantaggio di questo criterio è che esso non fa intervenire direttamente il limite (come nella definizione di successione convergente) ma soltanto gli elementi della successione.

Conviene premettere la seguente

Definizione 3.4 - *Una successione $\{a_n\}$ si dice di Cauchy o fondamentale se:*

$$\forall \varepsilon > 0 \quad , \quad \exists N = N(\varepsilon) \ : \ |a_n - a_m| < \varepsilon \ per \ n, m \geq N \ . \tag{3.21}$$

Un primo risultato interessante che riguarda le successioni fondamentali è il seguente.

Lemma 3.15 - *Ogni successione fondamentale è limitata.*

Dimostrazione - Sia N_0 un numero tale che, se $m, n \geq N_0$ risulti $|a_n - a_m| < 1$; prendiamo $m = N_0$ cosicché risulta $|a_n - a_{N_0}| < 1$ per ogni $n \geq N_0$; cioè a partire da un certo posto in poi, tutti gli a_n cadono nell'intervallo $(a_{N_0} - 1, a_{N_0} + 1)$. Posto allora $M = \max_{n<N_0} |a_n| + |a_{N_0}| + 1$ risulterà, per ogni n, $|a_n| < M$. □

Il seguente risultato occupa, come vedremo, una posizione di particolare rilievo nell'Analisi matematica.

▪ **Teorema 3.16** - (Criterio di convergenza di Cauchy). *Una successione $\{a_n\}$ è convergente se e solo se è fondamentale.*

Dimostrazione -
i) Convergente ⇒ fondamentale. Se l è il limite di $\{a_n\}$, per ogni $\varepsilon > 0$ esiste un $N = N(\varepsilon)$ tale che $|a_n - l| < \varepsilon/2$ se $n \geq N$; se anche $m \geq N$ si ha pure $|a_m - l| < \varepsilon/2$ e perciò $|a_n - a_m| = |a_n - l + l - a_m| \leq |a_n - l| + |a_m - l| < \varepsilon$.

ii) Fondamentale ⇒ convergente. Se $\{a_n\}$ è fondamentale, allora è limitata (Lemma 3.15); pertanto è possibile estrarre da essa una sottosuccessione $\{a_{k_n}\}$ convergente (per esempio, convergente al massimo limite). Sia l il limite di questa sottosuccessione. Fissato $\varepsilon > 0$, esistono N_1 e N_2 tali che:

$$|a_{k_n} - l| < \frac{\varepsilon}{2} \quad \text{per } n \geq N_1$$

$$|a_n - a_m| < \frac{\varepsilon}{2} \quad \text{per } n, m \geq N_2 \ .$$

In particolare , per $m = k_n$ sarà

$$|a_n - a_{k_n}| < \frac{\varepsilon}{2} \quad \text{per } n \geq N_2 \ .$$

Prendendo $n \geq \max(N_1, N_2)$ avremo

$$|a_n - l| \leq |a_n - a_{k_n}| + |a_{k_n} - l| \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

cosicché tutta la successione converge a l. □

Il criterio di Cauchy è una proprietà strettamente collegata con le proprietà di $\mathbb{R}$, in particolare con la proprietà di continuità $\mathcal{R}_4$; si può dimostrare infatti che la $\mathcal{R}_4$ è equivalente al criterio di Cauchy unito al principio di Archimede. Si sarà notato inoltre come la teoria delle successioni ha notevoli relazioni con la topologia di $\mathbb{R}$.

Queste osservazioni verranno riprese e illustrate più chiaramente nella sezione 4, trattando delle successioni a valori in $\mathbb{R}^n$.

Infine osserviamo che il criterio di Cauchy può essere esteso anche alle funzioni. Ci limitiamo ad enunciarlo; la dimostrazione ripete gli argomenti già utilizzati nel caso delle successioni.

Proposizione 3.17 - *Sia* $f : X \to \mathbb{R}$ *e sia* $x_0 \in \mathbb{R}^*$ *oppure* $\dot{\mathbb{R}}$ *un punto di accumulazione per* X. *Condizione necessaria e sufficiente affinché esista finito* $\lim_{x \to x_0} f(x)$ *è che, fissato comunque* $\varepsilon > 0$, *si possa trovare un intorno* U *di* x_0 *(dipendente in generale da* ε*), tale che: per qualunque coppia di punti* $x', x'' \in X \cap U \setminus \{x_0\}$ *si abbia*

$$|f(x') - f(x'')| < \varepsilon .$$

$*$ 3.6 Frazioni continue

Un numero reale irrazionale $\alpha = a_0 {\cdot} a_1 a_2 a_3 \cdots$, come allineamento decimale è approssimato in modo naturale dalla successione di numeri razionali costituita dalle sue troncate n–esime $\alpha^{(n)} = a_0 {\cdot} a_1 a_2 a_3 \cdots a_n$.

Un altro modo di approssimare numeri irrazionali mediante successioni di razionali è costituito dallo sviluppo in frazioni continue.

Cominciamo a considerare un numero razionale $p = \dfrac{n_0}{n_1}$, dove $n_0 \geq n_1$, n_0 ed n_1 interi positivi.

Eseguendo la divisione di n_0 per n_1 otteniamo un quoziente $q_1 = [n_0/n_1]$ ed un resto n_2 tale che $n_2 < n_1$ e $n_2 = n_0 - q_1 n_1$.

L'ultima relazione si può scrivere nella forma

$$\frac{n_0}{n_1} = q_1 + \frac{1}{n_1/n_2} .$$

Eseguiamo ora su n_1 ed n_2 (si noti che $n_2 < n_1$) le stesse operazioni; otteniamo

$$\frac{n_0}{n_1} = q_1 + \cfrac{1}{q_2 + \cfrac{1}{n_2/n_3}}$$

dove

$$q_2 = \left[\frac{n_1}{n_2}\right] \text{ e } n_3 = n_1 - n_2 q_2 .$$

In generale definiamo per $j \geq 1$

$$q_j = \left[\frac{n_{j-1}}{n_j}\right] \text{ e } n_{j+1} = n_{j-1} - n_j q_j \qquad (n_j < n_{j-1}) .$$

Osserviamo che l'ultima relazione si può scrivere nella forma

$$\frac{n_{j+1}}{n_j} = \frac{1}{\dfrac{n_j}{n_{j-1}} - q_j} \,. \tag{3.22}$$

Poiché $n_j < n_{j-1}$ il procedimento si arresta dopo k passi, dove k è il primo intero tale che $n_{k+1} = 0$.

Si noti che, allora, n_k è il massimo comun divisore tra n_0 ed n_1 (dimostrarlo per esercizio).

Mediante i quozienti $q_1, \cdots, q_k$ possiamo ricostruire n_0/n_1 con la formula seguente

$$\frac{n_0}{n_1} = q_1 + \cfrac{1}{q_2 + \cfrac{1}{q_3 + \cfrac{1}{\cdots \cfrac{}{q_{k-1} + \cfrac{1}{q_k}}}}} \tag{3.23}$$

L'espressione a destra della (3.23) si chiama sviluppo in frazione continua (limitata) del numero n_0/n_1 e viene indicata col simbolo $[q_1, q_2, \cdots, q_k]$.

Dunque

$$\frac{n_0}{n_1} = [q_1, \cdots, q_k] \,.$$

Un'altra notazione in uso è

$$q_1 + \frac{1}{q_2 +} \frac{1}{q_3 +} \cdots \frac{1}{+ q_k} \,.$$

Ad esempio:

$$\frac{15}{9} = [1, 1, 2], \quad \frac{173}{9} = [19, 4, 2] \,.$$

Se $n_0 < n_1$ si possono ripetere gli stessi discorsi, osservando che dopo il primo passo si ritrova la divisione tra n_1 ed n_0. Dunque se $n_1/n_0 = [q_1, \cdots, q_k]$ si avrà $n_0/n_1 = [0, q_1, \cdots, q_k]$.

Sia ora $\alpha > 0$ un numero irrazionale. Definiamo ricorsivamente: $\alpha_1 = \alpha$, $q_1 = [\alpha_1]$ e, per $j \geq 2$ (cfr. la (3.22))

$$\alpha_j = \frac{1}{\alpha_{j-1} - q_{j-1}} \quad \text{e} \quad q_j = [\alpha_j] \,. \tag{3.24}$$

Osserviamo che ogni differenza $\alpha_{j-1} - q_{j-1}$ è un numero irrazionale tra 0 e 1; dunque $\alpha_j > 1 \; \forall j$, ed il procedimento non si arresta mai. Al numero α si può

allora associare una successione di razionali $\{r_n\}$, detta successione delle *ridotte* (di α) definita da

$$r_n := [q_1, q_2, \cdots, q_n] .$$

Poiché, come vedremo subito, $r_n \to \alpha$, possiamo scrivere la formula

$$\alpha = [q_1, q_2, \cdots, q_k, \cdots]$$

che chiamiamo *sviluppo in frazione continua illimitata* di α.

Ad esempio si trova

$$\begin{aligned}
\sqrt{3} &= [1,1,2,1,2,1,2,1,2,\cdots] \\
\sqrt{7} &= [2,1,1,1,4,1,1,1,4,\cdots] \\
e &= [2,1,2,1,1,4,1,1,6,1,1,8,1,1,10,1,1,8,1,6,\cdots] \\
\pi &= [3,7,15,1,292,1,1,1,2,1,4,1,2,9,1,2,\cdots] .
\end{aligned}$$

Lo studente è invitato a verificare i precedenti sviluppi (ad esempio mediante una calcolatrice tascabile programmabile).

Vale il seguente teorema.

▪ **Teorema 3.18** - *Sia α irrazionale positivo ed $\{r_n\}$ la successione delle sue ridotte n–esime. Valgono le seguenti proprietà:*

i) $r_{2n-1} < \alpha < r_{2n}$ (cioè le ridotte di indice pari (dispari) approssimano per eccesso (difetto) α).

ii) $r_n \to \alpha$; più precisamente, detta $\{F_n\}$ la successione di Fibonacci definita ricorsivamente da $F_0 = 0, F_1 = 1, F_{n+1} = F_n + F_{n-1}$, si ha

$$|r_n - \alpha| < \frac{1}{F_n \cdot F_{n+1}} .$$

Per la dimostrazione del teorema ci serviamo del seguente lemma, riguardante numeratori e denominatori delle ridotte.

Lemma 3.19 - *Poniamo $r_n = a_n/b_n$ con a_n e b_n interi positivi primi tra loro. Si ha:*

i) a_n e b_n sono definiti ricorsivamente dalle seguenti formule:

$$\begin{aligned}
&a_0 = 1, \ a_1 = q_1; \quad a_{n+1} = a_n q_{n+1} + a_{n-1} \\
&b_0 = 0, \ b_1 = 1 \ ; \quad b_{n+1} = b_n q_{n+1} + b_{n-1}
\end{aligned}$$

ii) $a_{n-1} b_n - a_n b_{n+1} = (-1)^n$.

Cominciamo col dimostrare il teorema.

i) Dalla (3.24) si ottiene subito che

$$\alpha = \alpha_1 = q_1 + \frac{1}{\alpha_2} = q_1 + \cfrac{1}{q_2 + \cfrac{1}{\alpha_3}} = q_1 + \cfrac{1}{q_2 + \cfrac{1}{q_3 + \cfrac{1}{\alpha_4}}} \cdots ,$$

ovvero: $\alpha = [q_1, q_2, \cdots, q_{n-1}, \alpha_n]$ ovvero: sostituendo α_n al posto di q_n nella ridotta r_n si ottiene α.

Essendo, per la *i*) del lemma,

$$r_n = \frac{a_n}{b_n} = \frac{a_{n-1}\; q_n + a_{n-2}}{b_{n-1}\; q_n + b_{n-2}}$$

si ottiene

$$\alpha = \frac{a_{n-1}\alpha_n + a_{n-2}}{b_{n-1}\alpha_n + b_{n-2}} .$$

Dunque

$$\alpha - r_{n-1} = \frac{a_{n-1}\alpha_n + a_{n-2}}{b_{n-1}\alpha_n + b_{n-2}} - \frac{a_{n-1}}{b_{n-1}} = \frac{a_{n-2}b_{n-1} - a_{n-1}b_{n-2}}{b_{n-1}(b_{n-1}\alpha_n + b_{n-2})} =$$

$$\text{(per la } ii\text{) del lemma 3.19)} = \frac{(-1)^{n-1}}{b_{n-1}(b_{n-1}\alpha_n + b_{n-2})} .$$

Si vede allora che, se n è pari, $\alpha - r_{n-1} < 0$, ment..., se n è dispari, $\alpha - r_{n-1} > 0$, da cui la tesi.

Per la *ii*) osserviamo che, dalla *i*) del lemma, essendo $q_n \geq 1$, si ottiene $b_n \geq F_n$, $\forall n \geq 1$.

D'altra parte, usando la *i*) del teorema e poi la *ii*) del lemn a, si deduce che

$$|\alpha - r_n| < |r_{n+1} - r_n| = \left|\frac{a_{n+1}}{b_{n+1}} - \frac{a_n}{b_n}\right| = \frac{1}{b_n b_{n+1}} \leq \frac{1}{F_n F_{n+1}} .$$

Ricordando che $F_n \to +\infty$ si ricava la tesi. □

Rimane da dimostrare il lemma.

La *ii*) si ricava dalla *i*): moltiplicando per b_n la relazione di ricorrenza per a_{n+1} e per a_n quella per b_{n+1} e poi sottraendo membro a membro si ottiene

$$a_{n+1}b_n - a_n b_{n+1} = -(a_n b_{n-1} - a_{n-1}b_n) .$$

La *ii*) si deduce osservando che $a_1 b_0 - a_0 b_1 = -1$.

Incidentalmente la *ii*) implica che a_n e b_n sono effettivamente primi tra loro. Dimostriamo ora la *i*) per induzione, partendo da $n = 2$.

Abbiamo

$$r_1 = \frac{a_2}{b_2} = q_1 + \frac{1}{q_2} = \frac{q_1 q_2 + 1}{q_2} = \frac{a_1 q_2 + a_0}{b_1 q_2 + b_0}$$

e quindi $a_2 = a_1 q_2 + a_0$ e $b_2 = b_1 q_2 + b_0$.

Supponiamo ora che le formule siano vere per l'intero n, cioè $r_n = \dfrac{a_{n-1}q_n + a_{n-2}}{b_{n-1}q_n + b_{n-2}}$. Allora

$$r_{n+1} = \frac{a_{n-1}(q_n + 1/q_{n+1}) + a_{n-2}}{b_{n-1}(q_n + 1/q_{n+1}) + b_{n-2}} = \frac{a_{n-1}/q_{n+1} + a_{n-1}q_n + a_{n-2}}{b_{n-1}/q_{n+1} + b_{n-1}q_n + b_{n-2}} =$$

(usando ancora l'ipotesi di ricorrenza)

$$= \quad \frac{a_{n-1}/q_{n+1} + a_n}{b_{n-1}/q_{n+1} - b_n} = \frac{a_n q_{n+1} + a_{n-1}}{b_n q_{n+1} + b_{n-1}} \,. \ \square$$

Concludiamo il nostro breve cenno alle frazioni continue con qualche osservazione.

La prima riguarda la differenza tra gli sviluppi di $\sqrt{7}$ e $\sqrt{3}$ e quelli di e o π.

Mentre nei secondi la sequenza dei quozienti q_1 appare piuttosto "selvaggia" nei primi due si susseguono gruppi di quozienti, con periodicità.

Infatti si dice che gli sviluppi di $\sqrt{3}$ e $\sqrt{7}$ sono periodici e si scrive

$$\sqrt{3} = [1, \overline{1, 2}] \quad , \quad \sqrt{7} = [2, \overline{1, 1, 1, 4}] \,.$$

Proviamo con lo sviluppo di $\sqrt[3]{3}$: si trova

$$\sqrt[3]{3} = [1, 2, 3, 1, 4, 1, 5, 1, 1, 6, \cdots] \,.$$

Non sembra ci sia periodicità. Infatti vale il seguente teorema dovuto a Lagrange:

Lo sviluppo in frazione continua di un numero razionale α è periodico se e solo se α è soluzione di un'equazione di 2° grado con coefficienti interi.

La seconda osservazione riguarda la successione di Fibonacci $F_0 = 0$, $F_1 = 1$, $F_{n+1} = F_n + F_{n-1}$ che compare nel teorema 3.14 e che, del resto, avevamo già avuto modo di incontrare.

Dalla relazione di ricorrenza si ottiene dividendo per F_n, $n \geq 2$

$$\frac{F_{n+1}}{F_n} = 1 + \frac{1}{\dfrac{F_n}{F_{n-1}}}$$

che, ponendo $s_n = F_n/F_{n-1}$ possiamo scrivere nella forma

$$s_{n+1} = 1 + \frac{1}{s_n}, \quad \text{con } s_1 = 1 \,. \tag{3.25}$$

Possiamo seguire graficamente l'andamento della successione s_n considerando la funzione $f(x) = 1 + 1/x$ e osservando che $s_{n+1} = f(s_n)$.

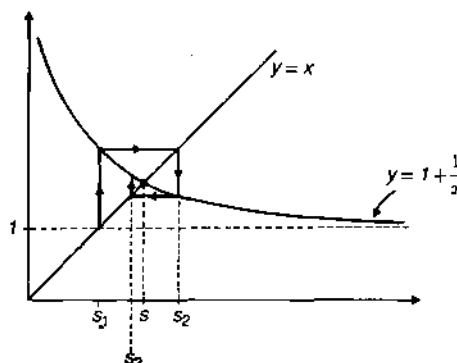



Fig. 4.16

Essendo f decrescente si ha (il lettore lo verifichi analiticamente)

$$s_1 < s_3 < s_5 < \cdots < s_{2n-1} < s_{2n+1} < \cdots < s_2$$

mentre

$$s_2 > s_4 > s_6 > \cdots > s_{2n-2} > s_{2n} > \cdots > s_1 \, .$$

Dunque le successioni $\{s_{2n+1}\}$ ed $\{s_{2n}\}$ ammettono limite finito, l_- ed l_+ rispettivamente.

Poiché da (3.25) si ha:

$$s_{n+1} = 1 + \cfrac{1}{1 + \cfrac{1}{s_{n-1}}}$$

le due successioni soddisfano la stessa relazione di ricorrenza. Passando al limite in essa si ottiene che l_+ ed l_- devono coincidere con la soluzione positiva dell'equazione

$$l^2 - l - 1 = 0 \, .$$

Tale soluzione è l'ascissa s del punto di intersezione tra f e la bisettrice del 1° quadrante; si ha

$$s = \frac{1+\sqrt{5}}{2} \quad \text{(il “rapporto aureo” dei Greci)} \, .$$

In conclusione, se F_n è la successione di Fibonacci

$$\frac{F_n}{F_{n-1}} \to \frac{1+\sqrt{5}}{2} \, .$$

Ora, lo sviluppo in frazioni continue di $\dfrac{1+\sqrt{5}}{2}$ è $[1,\bar{1}]$ e cioè tutti i numeri q_k sono uguali a 1.

Dalla *i*) del lemma otteniamo allora che $b_n = F_n$ e $a_n = F_{n+1}$ ovvero che i rapporti F_{n+1}/F_n coincidono con le ridotte n–esime del rapporto aureo.

Il Fibonacci (Leonardo Pisano, figlio di Bonaccio) era stato condotto alla successione che porterà il suo nome da un problema di ... allevamento.

"Quante coppie di *conigli* possono essere prodotte in un anno a partire da una sola coppia?"

Le ipotesi (sperimentali) erano:

1) ogni coppia ne genera un'altra ogni mese;

2) ogni nuova coppia diventa fertile dopo un mese;

3) i conigli non muoiono mai.

Il numero di coppie dopo n–mesi è esattamente F_n.

4. LIMITI IN $\mathbb{C}$. LIMITI IN $\mathbb{R}^n$

4.1 Funzioni da $\mathbb{R}^n$ in $\mathbb{R}^m$ e loro limiti

Sia $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ un punto di $\mathbb{R}^n$ e $X \subseteq \mathbb{R}^n$; le scritture

$$f : \mathbb{R}^n \supseteq X \to \mathbb{R} \ , \ \mathbf{x} \mapsto f(\mathbf{x}) \ \text{ oppure } \ y = f(\mathbf{x})$$

indicano una funzione reale di n variabili reali definita su X.

Esempi

4.1. $\mathbf{x} \mapsto \|\mathbf{x}\| = \sqrt{x_1^2 + \cdots + x_n^2}$ è la funzione che associa, ad ogni punto di $\mathbb{R}^n$, la sua distanza dall'origine.

4.2. $\mathbf{x} \mapsto x_k$ (k fisso, $1 \le k \le n$) è la funzione che associa, ad ogni punto di $\mathbb{R}^n$, la sua coordinata k–esima.

4.3. Sia $\mathbf{a} = (a_1, a_2, \cdots, a_n)$. La funzione $\mathbf{x} \mapsto < \mathbf{a}, \mathbf{x} > = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$ rappresenta un piano passante per l'origine.

Poiché l'immagine $f(X)$ di tali funzioni è contenuta in un insieme ordinato, lo spazio $\mathbb{R}$, immediatamente si trasportano a queste funzioni le definizioni di:

funzione positiva (negativa)	: $f(X) \subseteq \mathbb{R}_+$ $(\mathbb{R}_-)$
funzione limitata superiormente	: $f(X) \subseteq \mathbb{R}$ è limitata superiormente
funzione limitata inferiormente	: $f(X) \subseteq \mathbb{R}$ è limitata inferiormente
estremo superiore (inferiore) di f	: estremo superiore (inferiore) di $f(X)$
massimo (minimo) di f	: massimo (minimo) di $f(X)$.

Così anche la definizione di punti di massimo (minimo) globale e locale si estende senza difficoltà.

Esempi

4.4. La funzione caratteristica di un insieme $E \subset \mathbb{R}^n$, cioè l'applicazione $\mathbf{x} \mapsto f(\mathbf{x}) = 1$ se $\mathbf{x} \in E$, $f(\mathbf{x}) = 0$ se $\mathbf{x} \notin E$, è non negativa, limitata; $\sup f = \max f = 1$, $\inf f = \min f = 0$; ogni punto $\mathbf{x} \in E$ è punto di massimo, ogni punto $\mathbf{x} \notin E$ è punto di minimo.

4.5. $\mathbf{x} \mapsto f(\mathbf{x}) = e^{x_1 + x_2 + \cdots + x_n}$ è positiva con $\inf f = 0$ ed illimitata superiormente.

4.6. $\mathbf{x} \mapsto e^{-\|\mathbf{x}\|^2}$ è positiva, limitata, $\inf f = 0$, $\sup f = \max f = 1$, non ammette minimo.

Definizione 4.1 - *Sia $f : \mathbb{R}^n \supseteq X \to \mathbb{R}$ e $\mathbf{x}^0 \in \dot{\mathbb{R}}^n$ un punto di accumulazione per X; sia inoltre l un elemento di $\mathbb{R}^*$ oppure di $\dot{\mathbb{R}}$. Diremo che: il limite di $f(\mathbf{x})$, per $\mathbf{x}$ che tende a $\mathbf{x}^0$, è l, e scriveremo:*

$$\lim_{\mathbf{x} \to \mathbf{x}^0} f(\mathbf{x}) = l \quad \textit{oppure} \quad f(\mathbf{x}) \to l \quad \textit{per} \quad \mathbf{x} \to \mathbf{x}^0$$

se, per ogni intorno V di l è possibile trovare un intorno U di $\mathbf{x}^0$ per cui

$$f(\mathbf{x}) \in V \quad \textit{se} \quad \mathbf{x}^0 \neq \mathbf{x} \in U \cap X \, .$$

Come si vede nulla è cambiato rispetto alla Definizione 2.2 (*).

Come abbiamo già fatto in 2.1, la definizione può essere resa più esplicita nei vari casi particolari. Per esempio, se $\mathbf{x}^0 \in \mathbb{R}^n$ e $l \in \mathbb{R}$ diremo: $f(\mathbf{x}) \to l$ per $\mathbf{x} \to \mathbf{x}^0$ se, fissato $\varepsilon > 0$, si può trovare un $\delta = \delta(\varepsilon) > 0$ tale che

$$\|f(\mathbf{x}) - l\| < \varepsilon \quad \text{se} \quad \mathbf{x} \in X \text{ e } 0 < \|\mathbf{x} - \mathbf{x}^0\| < \delta \, .$$

(*) Osservazione sulla notazione: dovendo usare degli indici per indicare diversi punti $\mathbf{x}$ di $\mathbb{R}^n$ sceglieremo di volta in volta di metterli in basso: $\mathbf{x}_0$, $\{\mathbf{x}_k\}$ oppure in alto: $\mathbf{x}^0$, $\{\mathbf{x}^k\}$ secondo la comodità di scrittura. L'indice in alto sarà preferito quando dovranno essere indicate le componenti del vettore; $\mathbf{x}^k = (x_1^k, x_2^k, \cdots, x_n^k)$.

Si osservi che la condizione $||\mathbf{x}-\mathbf{x}^0|| < \delta$, cioè $\sqrt{(x_1 - x_1^0)^2 + \cdots + (x_n - x_n^0)^2} <$ $< \delta$ comporta che, per *ogni* $k = 1, 2, \cdots, n$ risulti $|x_k - x_k^0| < \delta$.

Poiché l'insieme di definizione X di f non è ordinato, non si può parlare di limite destro e sinistro; ma poiché l'immagine di f è contenuta in $\mathbb{R}$, spazio ordinato, si può ancora parlare di limite per difetto e per eccesso. Così pure si estendono, oltre al teorema di unicità del limite, tutti i teoremi del calcolo dei limiti, sia quelli che riguardano l'ordinamento (dell'insieme immagine): permanenza del segno e confronto, sia quelli che riguardano le operazioni algebriche; lo studente è invitato a ripetere qualcuna delle semplici dimostrazioni come esercizio.

Osservazione 4.1 - Anche se la struttura della definizione di limite rimane inalterata, il calcolo dei limiti per funzioni di più variabili è in generale tecnicamente piuttosto complicato.

Va infatti sottolineato che mentre in $\mathbb{R}^*$ (o $\dot{\mathbb{R}}$), x tende ad un punto x_0 lungo una direzione fissata (con due possibili versi) in $\dot{\mathbb{R}}^n$ $\mathbf{x}$ può tendere ad un punto $\mathbf{x}^0$ con gli n gradi di libertà che lo spazio gli consente.

La definizione richiede che $f(\mathbf{x})$ tenda a l *indipendentemente* da *come* $\mathbf{x}$ si avvicina ad $\mathbf{x}^0$; l'unica cosa che conta è che la *distanza* (euclidea) di $\mathbf{x}$ da $\mathbf{x}^0$ tenda a zero.

Ad esempio in $\mathbb{R}^2$, $\mathbf{x}$ può tendere all'origine $\mathbf{0}$ muovendosi lungo rette, parabole, spirali o qualunque altra "curva" (fig. 4.17).

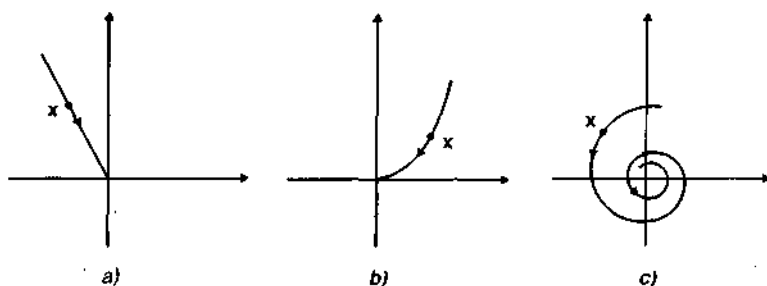



Fig. 4.17

Esempi

4.7. $\lim_{\mathbf{x}\to\infty} \frac{1}{||\mathbf{x}||} = 0$. Infatti, fissato $\varepsilon > 0$, si trova $k = k(\varepsilon)$ tale che: $1/||\mathbf{x}|| < \varepsilon$ se $||\mathbf{x}|| > k$; basterà prendere $k = 1/\varepsilon$.

4.8. $\lim_{\mathbf{x}\to 0} \frac{x_1}{||\mathbf{x}||}$ non esiste. Mostriamolo per $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$. Basta osservare che la funzione $x_1 \Big/ \sqrt{x_1^2 + x_2^2}$ vale 0 nei punti $(0\,, x_2)$ con $x_2 \neq 0$, vale $+1$

nei punti $(x_1, 0)$ con $x_1 > 0$ e vale -1 nei punti $(x_1, 0)$ con $x_1 < 0$, cosicché se facciamo tendere $\mathbf{x}$ a $\mathbf{0}$ lungo l'asse x_2 il "limite" risulta 0, mentre se $\mathbf{x}$ tende a $\mathbf{0}$ lungo l'asse x_1 per valori postivi di x_1 il "limite" è 1, per valori negativi di x_1 il "limite" è -1.

In $\mathbb{R}^2$ il calcolo dei limiti può essere facilitato con l'uso delle coordinate polari. Indichiamo con x, y le coordinate cartesiane di un punto di $\mathbb{R}^2$, e con ρ, θ le sue coordinate polari. Sia $(x_0, y_0) \in \mathbb{R}^2$. Poniamo

$$x = x_0 + \rho\cos\theta \quad , \quad y = y_0 + \rho\sin\theta \; .$$

Si osserva che $(x, y) \to (x_0, y_0)$ equivale a $\rho \to 0_+$, *indipendentemente* da θ. Precisamente si avrà: $\lim\limits_{(x,y)\to(x_0,y_0)} f(x,y) = l$ se e solo se, posto

$$\tilde{f}(\rho, \theta) = f(x_0 + \rho\cos\theta, y_0 + \rho\sin\theta)$$

risulta che, per ogni fissato $\varepsilon > 0$, esiste $\delta = \delta(\varepsilon)$ (dipendente da ε, ma non da θ) tale che:

$$|\tilde{f}(\rho, \theta) - l| < \varepsilon \quad \text{per} \quad 0 < \rho < \delta \quad \text{e} \quad \theta \in \mathbb{R} \; .$$

In questa situazione si dice che:

$$f(\rho, \theta) \to l \quad \text{per} \quad \rho \to 0 \quad \textit{uniformemente rispetto a } \theta \; .$$

Questo concetto (limite uniforme) sarà ripreso e sviluppato ancora meglio più avanti.

Esempi

4.9. Si voglia calcolare $\lim\limits_{(x,y)\to(0,0)} \dfrac{x^3}{x^2 + y^2}$. Posto $x = \rho\cos\theta$, $y = \rho\sin\theta$ il limite proposto diventa: $\lim\limits_{\rho\to 0_+} \rho(\cos\theta)^3$; essendo $|\cos\theta| \le 1$ tale limite è 0, uniformemente rispetto a θ.

4.10. $\lim\limits_{(x,y)\to(0,0)} \dfrac{x+y}{x-y} = \lim\limits_{\rho\to 0_+} \dfrac{\cos\theta + \sin\theta}{\cos\theta - \sin\theta} = \dfrac{\cos\theta + \sin\theta}{\cos\theta - \sin\theta}$, funzione di θ. Perciò il limite proposto non esiste.

Sia ora $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ un punto di $\mathbb{R}^n$ e $\mathbf{y} = (y_1, y_2, \cdots, y_m)$ un punto di $\mathbb{R}^m$. Le scritture

$$\mathbf{f} : \mathbb{R}^n \supseteq X \to \mathbb{R}^m \;\; , \;\; \mathbf{x} \mapsto \mathbf{f}(\mathbf{x}) \quad \text{oppure} \quad \mathbf{y} = \mathbf{f}(\mathbf{x})$$

indicano una applicazione da $\mathbb{R}^n$ in $\mathbb{R}^m$ o, come si dice, una *funzione a valori vettoriali*. Indicando esplicitamente le m componenti di $\mathbf{f}, \mathbf{f} = (f_1, f_2, \cdots, f_m)$

possiamo riguardare una tale applicazione come l'insieme di m funzioni reali di n variabili

$$\begin{aligned} f_1 &: \mathbf{x} \mapsto f_1(\mathbf{x}) \\ f_2 &: \mathbf{x} \mapsto f_2(\mathbf{x}) \\ &\vdots \\ f_m &: \mathbf{x} \mapsto f_m(\mathbf{x}) . \end{aligned}$$

***Esempio* 4.11** - Siano $\mathbf{a}^1, \mathbf{a}^2, \cdots, \mathbf{a}^m$ m vettori di $\mathbb{R}^n$. Possiamo definire m funzioni da $\mathbb{R}^n$ in $\mathbb{R}$ ponendo

$$f_1 : \mathbf{x} \mapsto < \mathbf{a}^1, \mathbf{x} > \ , \ f_2 : \mathbf{x} \mapsto < \mathbf{a}^2, \mathbf{x} >, \cdots, f_m : \mathbf{x} \mapsto < \mathbf{a}^m, \mathbf{x} > \ .$$

Se indichiamo con A la matrice $m \times n$ ottenuta incolonnando i vettori $\mathbf{a}^1, \cdots, \mathbf{a}^m$

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}^1 \\ \mathbf{a}^2 \\ \vdots \\ \mathbf{a}^m \end{pmatrix} = \begin{pmatrix} a_1^1 & a_2^1 & \dots & a_n^1 \\ a_1^2 & a_2^2 & \dots & a_n^2 \\ \vdots & \vdots & & \\ a_1^m & a_2^m & \dots & a_n^m \end{pmatrix}$$

possiamo sinteticamente rappresentare le m funzioni precedenti come una funzione $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ così definita

$$\mathbf{f} : \mathbf{y} = \mathbf{A}\mathbf{x}$$

dove $\mathbf{Ax}$ rappresenta il vettore di $\mathbb{R}^m$ ottenuto eseguendo il prodotto righe per colonne della matrice $\mathbf{A}$ $(m \times n)$ per il vettore $\mathbf{x}$ $(n \times 1)$.

Le applicazioni di $\mathbb{R}^2$ in $\mathbb{R}^2$ sono particolarmente interessanti poiché possono essere identificate come le funzioni complesse di variabile complessa, cioè le applicazioni da $\mathbb{C}$ in $\mathbb{C}$: infatti, come abbiamo già osservato, dal punto di vista insiemistico $\mathbb{C}$ e $\mathbb{R}^2$ coincidono.

Per le funzioni di variabile complessa si usano generalmente le seguenti scritture:

$$f : \mathbb{C} \supseteq X \to \mathbb{C} \ , \ z \mapsto f(z) \text{ oppure } w = f(z) .$$

Scrivendo $z = x + iy$ e $w = u + iv$ (x, y, u, v reali) e separando parte reale e immaginaria nella precedente scrittura, possiamo rappresentare la funzione $z \to f(z)$ come la coppia di funzioni reali

$$(x, y) \mapsto u(x, y)$$

$$(x, y) \mapsto v(x, y)$$

cioè come un'applicazione da $\mathbb{R}^2$ in $\mathbb{R}^2$.

***Esempio* 4.12** - La funzione complessa $w = z^2$ cioè $u+iv = (x^2-y^2)+i(2xy)$ è equivalente alla coppia di funzioni reali

$$\begin{cases} u(x,y) = x^2 - y^2 \\ v(x,y) = 2xy \end{cases}$$

Poiché ora l'immagine $\mathbf{f}(X)$ dell'applicazione non è un insieme ordinato, non possiamo più parlare di estremo superiore, inferiore, di massimi e minimi; ma diremo ancora che $\mathbf{f}$ è *limitata* se $\mathbf{f}(X)$ è limitato, cioè se è contenuto in qualche intorno $B(\mathbf{0}, r)$ dell'origine in $\mathbb{R}^m$.

Anche la definizione di limite si ripete senza sostanziali varianti.

Definizione 4.2 - *Sia* $\mathbf{f} : \mathbb{R}^n \supseteq X \to \mathbb{R}^m$ *e sia* $\mathbf{x}^0 \in \dot{\mathbb{R}}^n$ *un punto di accumulazione per* X. *Sia inoltre* $\mathbf{l}$ *un elemento di* $\dot{\mathbb{R}}^m$. *La scrittura:*

$$\lim_{\mathbf{x}\to\mathbf{x}^0} \mathbf{f}(\mathbf{x}) = \mathbf{l}$$

significa che, per ogni intorno V *di* $\mathbf{l}$ *(in* $\dot{\mathbb{R}}^m$*) è possibile trovare un intorno* U *di* $\mathbf{x}^0$ *(in* $\dot{\mathbb{R}}^n$*) per cui succede che*

$$\mathbf{f}(\mathbf{x}) \in V \quad se \quad \mathbf{x}^0 \neq \mathbf{x} \in U \cap X \ .$$

Per le funzioni di variabile complessa la definizione è del tutto analoga:

$$\lim_{z\to z_0} f(z) = l$$

significa che, per ogni intorno V *di* l *(in* $\dot{\mathbb{C}}$*) è possibile trovare un intorno* U *di* z_0 *(in* $\dot{\mathbb{C}}$*) per cui succede che*

$$f(z) \in V \quad se \quad z_0 \neq z \in U \cap X \ .$$

Poiché la topologia adottata per $\dot{\mathbb{C}}$ è la stessa adottata per $\dot{\mathbb{R}}^2$ (gli intorni sono gli stessi!) la definizione di limite per le funzioni di variabile complessa è un caso particolare della definizione 4.2.

Si osservi infine che, se $f_1, f_2, \cdots, f_m$ sono le componenti di $\mathbf{f}$ e $l_1, l_2, \cdots, l_m$ le componenti di $\mathbf{l}$ la scrittura $\lim_{\mathbf{x}\to\mathbf{x}^0} \mathbf{f}(\mathbf{x}) = \mathbf{l}$ è equivalente a

$$\lim_{\mathbf{x}\to\mathbf{x}^0} f_1(\mathbf{x}) = l_1$$

$$\vdots$$

$$\lim_{\mathbf{x}\to\mathbf{x}^0} f_m(\mathbf{x}) = l_m$$

Esercizi

1. Dimostrare che:

a) $\lim\limits_{(x,y)\to(0,0)} xe^{-y^2/x^2} = 0$

b) $\lim\limits_{(x,y)\to(0,0)} xe^{-y^2/x}$ non esiste.

2. Sia $f(x,y) \to l \in \mathbb{R}$ per $(x,y) \to (x_0, y_0)$. Supponiamo che i limiti (unidimensionali)

$$\lim_{x\to x_0} f(x,y) \quad \text{e} \quad \lim_{y\to y_0} f(x,y)$$

esistano finiti, il primo per ogni y in un intorno di y_0, e il secondo per ogni x in un intorno di x_0. Dimostrare che

$$\lim_{y\to y_0} \lim_{x\to x_0} f(x,y) = \lim_{x\to x_0} \lim_{y\to y_0} f(x,y) = l \ .$$

3. Sia $f(x,y) \to l \in \mathbb{R}$ per $(x,y) \to (x_0, y_0)$.
Dimostrare con un controesempio che in generale

$$\lim_{y\to y_0} \lim_{x\to x_0} f(x,y) \neq \lim_{x\to x_0} \lim_{y\to y_0} f(x,y)$$

[$f(x,y) = y\cos 1/x$.....]

4. Calcolare i seguenti limiti

a) $\lim\limits_{(x,y)\to(0,0)} (x^2+y^2)^\alpha \sin\dfrac{x}{y} \qquad (\alpha > 0)$

b) $\lim\limits_{(x,y)\to(0,0)} \dfrac{x^3 + xy^2}{(x^2+y^2)^{3/2}}$.

4.2 Successioni e topologia di $\mathbb{R}^n$

Come abbiamo fatto per le funzioni, così anche per le successioni a valori in $\mathbb{R}^n$ o in $\mathbb{C}$ (cioè applicazioni da $\mathbb{N}$ in $\mathbb{R}^n$ o da $\mathbb{N}$ in $\mathbb{C}$) possiamo ripetere la definizione di limite ed estendere le proprietà che non dipendono dall'ordinamento dell'insieme immagine. In particolare il teorema 3.6, che riconduce i limiti di funzioni a limiti di successioni, vale anche per funzioni da $\mathbb{R}^n$ in $\mathbb{R}^m$.

Esempi

4.13. La successione ($\mathbb{N} \to \mathbb{C}$) $\{i^h\}$ $(h = 0,1,2,\cdots)$ è limitata, ma non ammette limite. Questa successione si identifica con la successione ($\mathbb{N} \to \mathbb{R}^2$) i cui elementi sono:

$$(1,0),\ (0,1),\ (-1,0),\ (0,-1),\ (1,0),\ (0,1),\cdots .$$

4.14. La successione $(\mathbb{N} \to \mathbb{C})\,\{q^h\}$ con $q \in \mathbb{C}$ (*progressione geometrica*) presenta un comportamento analogo al caso reale (cfr. esempio 3.5) e cioè:

converge a zero	se $|q| < 1$
diverge a ∞	se $|q| > 1$

come subito si accerta.

Se $|q| = 1$ (come nell'esempio precedente) la successione non ha limite (tranne nel caso $q = 1$). Infatti, posto $q = \cos\theta + i\sin\theta$ (al variare di θ da 0 a 2π otteniamo tutti i numeri di modulo 1), abbiamo (per la formula di De Moivre): $q^h = \cos h\theta + i\sin h\theta$ ed è facile vedere che non esiste il limite per $h \to \infty$ (se $\theta \neq 0,\ 2\pi$).

In questo paragrafo vogliamo esaminare alcune importanti relazioni tra la teoria delle successioni e la topologia di $\mathbb{R}^n$, relazioni che sono state accennate nel caso particolare delle successioni a valori in $\mathbb{R}$, ma che risultano più significative nel caso più generale di successioni a valori in $\mathbb{R}^n$.

Successioni limitate e successioni convergenti. La proposizione 3.2 si estende senza difficoltà alle successioni a valori in $\mathbb{R}^n$.

Proposizione 4.1 - *Ogni successione a valori in $\mathbb{R}^n$ convergente è limitata.*

Come è noto, l'affermazione inversa è falsa.

Proposizione 4.2 - *Sia $E \subseteq \mathbb{R}^n$ e $\mathbf{y}$ un punto di accumulazione di E. Allora esiste una successione di punti $\mathbf{x}^h \in E$ distinti da $\mathbf{y}$ e convergente a $\mathbf{y}$.*

Dimostrazione - Ogni intorno $B(\mathbf{y}, r)$ di $\mathbf{y}$ conterrà un punto di E. Prendiamo $r = 1/h$ con $h = 1, 2, \ldots$, e scegliendo ogni volta un punto $\mathbf{x}^h \in B(\mathbf{y}, 1/h) \cap E, \mathbf{x}^h \neq \mathbf{y}$, otteniamo una successione $\{\mathbf{x}^h\}$ di punti di E chiaramente convergente a $\mathbf{y}$. □

Osservazione 4.2 - La proposizione 4.2 vale anche per insiemi $E \subseteq \dot{\mathbb{R}}^n$; cioè, se ∞ è un punto di accumulazione di E, esiste una successione di punti di E divergente a ∞. Di conseguenza, se E non è limitato, esiste una successione di punti di E divergente a ∞.

La proposizione 4.2 si collega col teorema di Bolzano–Weierstrass: questo afferma che ogni insieme infinito ammette almeno un punto di accumulazione (tale punto sta in $\mathbb{R}^n$ se l'insieme è limitato in $\mathbb{R}^n$, altrimenti sta in $\dot{\mathbb{R}}^n$). Una successione $\{\mathbf{x}^h\}$ o ha immagine finita (ma allora uno almeno dei suoi punti, diciamo $\mathbf{y}$, compare infinite volte nella successione) oppure infinita ed in questo caso essa ammette almeno un punto di accumulazione, diciamo ancora $\mathbf{y}$; in entrambi i casi esiste perciò una successione $\{\mathbf{y}^h\}$ di punti estratta dalla successione $\{\mathbf{x}^h\}$, tendente a $\mathbf{y}$. In particolare, se la successione $\{\mathbf{x}^h\}$ è limitata, $\mathbf{y} \in \mathbb{R}^n$ e la sottosuccessione è convergente.

Abbiamo così provato un sorta di inverso della proposizione 4.1, e cioè

Proposizione 4.3 - *Da ogni successione limitata in $\mathbb{R}^n$ si può estrarre una sottosuccessione convergente.*

Successioni e chiusura di un insieme. Per mezzo delle successioni possiamo caratterizzare un insieme chiuso. Cominciamo col dimostrare il seguente

■ **Teorema 4.4** - *Sia $E \subseteq \mathbb{R}^n$ e $\mathbf{y} \in \mathbb{R}^n$. Allora $\mathbf{y} \in \overline{E}$ se e solo se esiste una successione $\{\mathbf{x}^h\}$ di punti di E che tende a $\mathbf{y}$.*

Dimostrazione - *i*) Per ipotesi, $\mathbf{y} \in \overline{E} = E \cup E'$. Se $\mathbf{y} \in E$, la successione costante $\{\mathbf{x}^h\}$ con $\mathbf{x}^h = \mathbf{y}$ per ogni h, tende a $\mathbf{y}$. Sia allora $\mathbf{y} \in E'$ (e non in E): la tesi segue dalla Proposizione 4.2.

ii) Ora l'ipotesi è: esiste una successione ... e la tesi: $\mathbf{y} \in \overline{E}$. Neghiamo la tesi e mostriamo la falsità dell'ipotesi. Se $\mathbf{y} \in C\overline{E}$ (che è un aperto) esisterà un intorno $B(\mathbf{y}, r)$ di $\mathbf{y}$ nel quale non cadono punti di $\overline{E}$ e perciò nemmeno punti di E; allora non esiste alcuna successione $\{\mathbf{x}^h\}$ di punti di E convergente a $\mathbf{y}$. □

Corollario 4.5 - *Un insieme $E \subseteq \mathbb{R}^n$ è chiuso se e solo se, presa comunque una successione $\{\mathbf{x}^h\}$ a valori in E e convergente in $\mathbb{R}^n$, il limite della successione sta in E. In simboli:*

$$E \text{ chiuso} \Leftrightarrow \begin{cases} \mathbf{x}^h \in E \\ \mathbf{x}^h \to \mathbf{y} \end{cases} \Rightarrow \mathbf{y} \in E$$

Dimostrazione - *i*) L'implicazione $\Rightarrow$ segue immediatamente dal teorema 4.4.

ii) Per mostrare l'implicazione $\Leftarrow$, sia $\mathbf{y} \in E'$; allora esiste una successione $\{\mathbf{x}^h\}$ di punti di E convergente a $\mathbf{y}$; per l'ipotesi, $\mathbf{y} \in E$ e pertanto E è chiuso. □

Successioni e compattezza. In 3.2.3 abbiamo definito la nozione di compattezza per mezzo delle coperture. Abbiamo poi mostrato (teorema di Heine–Borel) che, in $\mathbb{R}^n$, sono compatti tutti e soli gli insiemi chiusi e limitati. La nozione di compattezza può essere introdotta anche per mezzo delle successioni.

Definizione 4.3 - *Un insieme $K \subset \mathbb{R}^n$ si dice compatto per successioni (o sequenzialmente compatto) se da ogni successione a valori in K si può estrarre una sottosuccessione convergente a un elemento di K.*

■ **Teorema 4.6** - *$K \subset \mathbb{R}^n$ è compatto se e solo se è sequenzialmente compatto.*

Dimostrazione - *i*) Se K è compatto e $\{\mathbf{x}^h\}$ è una successione a valori in K, da essa si può estrarre una sottosuccessione convergente a un punto $\mathbf{y} \in \mathbb{R}^n$ (perché K è limitato: Proposizione 4.3) e inoltre $\mathbf{y} \in K$ (perché K è chiuso: Corollario 4.5); perciò K è sequenzialmente compatto.

ii) Se K è sequenzialmente compatto, allora K è limitato: se così non fosse, esisterebbe una successione $\{\mathbf{x}^h\}$ di punti di K divergente a ∞; ogni successione estratta da questa

divergerebbe pure a ∞, per cui non sarebbe possibile estrarre da essa alcuna successione convergente.

iii) Se K è sequenzialmente compatto, allora K è chiuso: sia infatti $\mathbf{y} \in \overline{K}$; allora (teorema 4.4) esiste una successione $\{\mathbf{x}^h\}$ di punti di K convergente a $\mathbf{y}$. Per l'ipotesi, da questa successione posso estrarre una sottosuccessione convergente a un punto di K; ma ogni successione estratta dalla $\{\mathbf{x}^h\}$ converge a $\mathbf{y}$ per cui $\mathbf{y} \in K$. □

4.3 Il criterio di Cauchy

Ripetiamo la definizione di *successione di Cauchy* già introdotta per le successioni in $\mathbb{R}$.

Definizione 4.4 - *Sia $\{\mathbf{x}^h\}$ una successione a valori in $\mathbb{R}^n$ (o in $\mathbb{C}$). Se succede che, fissato $\varepsilon > 0$, esiste un $N = N(\varepsilon)$ per cui*

$$||\mathbf{x}^k - \mathbf{x}^h|| < \varepsilon \quad per \quad k, h \geq N \tag{4.1}$$

la successione si dirà di Cauchy o fondamentale.

La (4.1) si può anche scrivere

$$d(\mathbf{x}^k, \mathbf{x}^h) \to 0 \quad \text{per} \quad k, h \to +\infty \tag{4.2}$$

ricordando che $||\mathbf{x} - \mathbf{y}||$ esprime la distanza pitagorica dei punti $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.

Proposizione 4.7 - *Sia $\{\mathbf{x}^h\}$ una successione convergente in $\mathbb{R}^n$; allora è fondamentale.*

Dimostrazione - Essa ripete pari pari l'analoga svolta per le successioni in $\mathbb{R}$. Se $\mathbf{l} \in \mathbb{R}^n$ è il limite di $\{\mathbf{x}^h\}$ risulta, per la disuguaglianza triangolare

$$d(\mathbf{x}^k, \mathbf{x}^h) \leq d(\mathbf{x}^k, \mathbf{l}) + d(\mathbf{x}^h, \mathbf{l}) \tag{4.3}$$

e poiché i due addendi a secondo termine della (4.3) tendono a zero per $h, k \to +\infty$, così avviene anche del primo termine. □

Si noti che questa semplicissima dimostrazione fa uso soltanto della disuguaglianza triangolare, e non di proprietà particolari di $\mathbb{R}^n$. Essa resta valida perciò in uno spazio metrico qualsiasi (cfr. 3.14 per la definizione di spazio metrico) e non solo in $\mathbb{R}^n$.

Anche la proposizione inversa resta valida in $\mathbb{R}^n$.

Proposizione 4.8 - *Sia $\{\mathbf{x}^h\}$ una successione fondamentale in $\mathbb{R}^n$, allora essa è convergente.*

Ancora, la dimostrazione ricalca quella svolta in $\mathbb{R}$ e perciò non la ripetiamo (ma l'allievo lo faccia per esercizio). Ricordiamo però i tre passi della dimostrazione:

i) si mostra che $\{\mathbf{x}^h\}$, essendo fondamentale, è limitata (cfr. lemma 3.15)

ii) allora dalla $\{\mathbf{x}^h\}$ si può estrarre una sottosuccessione convergente (Proposizione 4.3)

iii) sfruttando ancora il fatto che $\{\mathbf{x}^h\}$ è fondamentale, si mostra che tutta la successione è convergente.

Mentre nei punti *i*) e *iii*) si fa uso sostanzialmente solo della disuguaglianza triangolare, nel punto *ii*) si ricorre al teorema di Bolzano–Weierstrass, che è una proprietà profonda e intimamente legata alla particolare struttura di $\mathbb{R}^n$.
Ciò giustifica l'introduzione della seguente nomenclatura.

Definizione 4.5 - *Uno spazio metrico X si dice* **completo** *se ogni successione fondamentale a valori in X è convergente.*

Mostriamo che esistono spazi metrici non completi.

***Esempio* 4.15** - L'insieme $\mathbb{Q}$ dei numeri razionali ha la struttura di spazio lineare metrico (sul campo $\mathbb{Q}$) con la solita definizione di distanza: $d(a,b) = |a-b|$.
Si consideri in $\mathbb{Q}$ la successione $\{(1+1/n)^n\}$ di numeri razionali; essa è *fondamentale* in $\mathbb{Q}$ (si può vederlo direttamente o anche si può dedurlo dal fatto che, essendo convergente in $\mathbb{R}$ necessariamente è fondamentale in $\mathbb{R}$ e quindi anche in $\mathbb{Q}$ perché abbiamo adottato in $\mathbb{Q}$ la stessa distanza presente in $\mathbb{R}$); essa *non* è *convergente* in $\mathbb{Q}$ poiché noi sappiamo che il suo limite, il numero e, non è razionale (ciò sarà dimostrato in 8.2.2) e un altro limite non può esservi, per l'unicità del limite!.
Ritroviamo così, da un altro punto di vista, una circostanza già più volte sottolineata: $\mathbb{R}$ è il completamento di $\mathbb{Q}$ nel senso che contiene $\mathbb{Q}$ e in più contiene i limiti delle successioni fondamentali di $\mathbb{Q}$ non convergenti in $\mathbb{Q}$.
Le precedenti proposizioni 4.7 e 4.8 si riassumono nel seguente teorema, fondamentale per l'Analisi matematica:

■ **Teorema 4.9** - *Gli spazi euclidei $\mathbb{R}^n$ e $\mathbb{C}^n$ sono completi.*

(La completezza di $\mathbb{C}^n$ segue da quella di $\mathbb{R}^n$ osservando che, sia dal punto di vista insiemistico che per la topologia, $\mathbb{C}^n$ è identificabile con $\mathbb{R}^{2n}$).
Il criterio di Cauchy, il teorema di Bolzano–Weierstrass, l'assioma di Dedekind, la proprietà dell'estremo superiore (proprietà $\mathcal{R}_4$ di 2.2.4), l'assioma di Cantor degli intervalli incapsulati (*), sono tutti collegati tra loro; si potrebbe dimostrare infatti

---

(*) Assioma di Cantor: ogni successione di intervalli chiusi $I_n = [a_n, b_n]$ tali che: $I_1 \supset I_2 \supset \cdots \supset I_n \supset \cdots$ ha intersezione non vuota.

la seguente catena di equivalenze:

$\mathcal{R}_4$

$\Updownarrow$

assioma di Dedekind

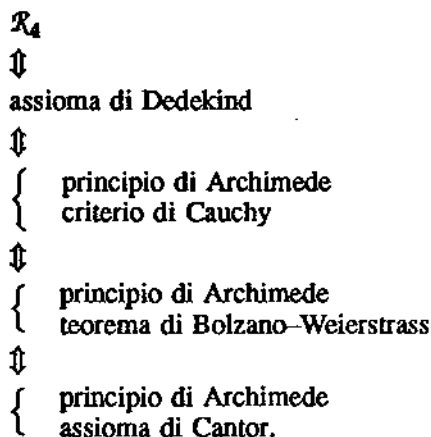
$\Updownarrow$

$\left\{\begin{array}{l}\text{principio di Archimede}\\ \text{criterio di Cauchy}\end{array}\right.$

$\Updownarrow$

$\left\{\begin{array}{l}\text{principio di Archimede}\\ \text{teorema di Bolzano–Weierstrass}\end{array}\right.$

$\Updownarrow$

$\left\{\begin{array}{l}\text{principio di Archimede}\\ \text{assioma di Cantor.}\end{array}\right.$

Concludiamo, accennando rapidamente alla *costruzione del campo dei numeri reali secondo Cantor*. Con un procedimento simile a quello che ci ha permesso di definire $\mathbb{Z}$ a partire da $\mathbb{N}$ e poi $\mathbb{Q}$ a partire da $\mathbb{Z}$, possiamo definire $\mathbb{R}$ a partire da $\mathbb{Q}$; ma non basterà utilizzare coppie o n–uple di razionali per definire un numero reale, bensì bisognerà utilizzare successioni.

Partiamo dunque da $\mathbb{Q}$ (insieme degli allineamenti decimali periodici strutturato come in 1.1.4) e costruiamo l'insieme C di tutte le successioni di Cauchy a valori in $\mathbb{Q}$.

In C introduciamo la seguente relazione di equivalenza:

$$(r) : \{a_n\} \approx \{b_n\} \;\Leftrightarrow\; \lim_{n\to+\infty}(a_n - b_n) = 0$$

(verificare che si tratta di una equivalenza!). Quindi si prende l'insieme quoziente C/r, che si indicherà con $\mathbb{R}$ e i cui elementi $[\{a_n\}]$ si diranno *numeri reali*.

Definiamo in $\mathbb{R}$ le *operazioni algebriche* di somma e prodotto:

$$[\{a_n\}] + [\{b_n\}] := [\{a_n + b_n\}]$$

$$[\{a_n\}] \cdot [\{b_n\}] := [\{a_n \cdot b_n\}]$$

(bisognerà verificare che, cambiando i rappresentanti nelle classi di equivalenza scritte a sinistra, non cambia la classe di equivalenza scritta a destra). Si mostra poi agevolmente che sono verificate le proprietà $\mathcal{R}_1$ e $\mathcal{R}_2$ di queste operazioni. Ovviamente il numero 0 è la classe di equivalenza della successione $\{a_n\}$ con $a_n = 0$ per ogni n (notare che di questa classe fanno parte, per esempio, le successioni $\{1/n\}$ e $\{1/n^2\}$, che sono equivalenti); il numero 1 è la classe di equivalenza della successione $\{b_n\}$ con $b_n = 1$ per ogni n; l'opposto di $\alpha = [\{a_n\}]$ sarà $-\alpha = [\{-a_n\}]$ e il reciproco $1/\alpha$ (se $\alpha \neq 0$) sarà $[\{1/a_n\}]$ (avendo eliminato dalla successione $\{a_n\}$ gli eventuali termini nulli).

Si introduce poi un *ordinamento totale* nel modo seguente: se $\alpha = [\{a_n\}]$ diremo che $\alpha \geq 0$ se nella classe di equivalenza che definisce α c'è una successione che risulti definitivamente ≥ 0; diremo che $\alpha \leq 0$ se $-\alpha \geq 0$; si mostra poi che, preso un qualsiasi $\alpha = [\{a_n\}]$ risulta $\alpha \geq 0$ oppure $\alpha \leq 0$. Dati due numeri, β e γ, diremo allora che $\beta \geq \gamma$ se $\beta - \gamma \geq 0$. Si mostra che le proprietà $\mathcal{R}_3$ sono verificate.

A questo punto possiamo *identificare* il numero razionale q con il numero reale definito dalla (classe di equivalenza della) successione costante $\{q\}$; la corrispondenza $q \to [\{q\}]$ conserva le operazioni algebriche e l'ordinamento sopra introdotti, cosicché possiamo dire che $\mathbb{Q}$ è un sottocampo di $\mathbb{R}$.

Infine si mostra che vale la proprietà $\mathcal{R}_4$.

Funzioni continue 5

Nel precedente capitolo, in cui è stata definita l'operazione di limite, abbiamo più volte sottolineato che il limite di una funzione $f(x)$, per $x \to x_0$, non è in alcun modo legato al valore che f assume in x_0, anche ammesso che f sia definita in x_0. È venuto ora il momento di esaminare una classe di funzioni $f(x)$ per le quali il limite per x tendente a un certo punto coincide col valore di f in quel punto. Queste funzioni sono fondamentali nell'Analisi matematica e prendono il nome di funzioni continue. In questo capitolo si studiano le proprietà di tali funzioni.

La continuità di una funzione, essendo legata alla nozione di limite, è una proprietà di carattere topologico; si potrebbe perciò affrontare subito lo studio della continuità delle funzioni da $\mathbb{R}^n$ in $\mathbb{R}^m$, dal momento che conosciamo la topologia di questi spazi e i limiti. Ma, per ragioni didattiche, e come già abbiamo fatto nel capitolo precedente, studieremo prima il caso più semplice delle funzioni da $\mathbb{R}$ in $\mathbb{R}$ (sez. 1); illustreremo poi (sez. 2) le proprietà fondamentali delle funzioni continue da $\mathbb{R}^n$ in $\mathbb{R}^m$ (e, in particolare, da $\mathbb{R}^n$ in $\mathbb{R}$ e da $\mathbb{C}$ in $\mathbb{C}$); infine (sez. 3) passeremo in rassegna le cosiddette "funzioni elementari" dell'Analisi.

1. FUNZIONI CONTINUE DA $\mathbb{R}$ IN $\mathbb{R}$

1.1 Definizione di continuità

Definizione 1.1 - *Sia $f : \mathbb{R} \supseteq X \to \mathbb{R}$. Se x_0 è un punto isolato di X, diremo senz'altro che f è continua in x_0.*

Se $x_0 \in X \cap X'$, cioè è un punto di accumulazione di X che appartiene a X, diremo che f è continua in x_0 se è verificata una delle seguenti circostanze, tra loro equivalenti:

i) per ogni intorno U di $f(x_0)$ esiste un intorno V di x_0 tale che risulti

$$f(x) \in U \quad \text{per ogni} \quad x \in V \cap X$$

ii) per ogni $\varepsilon > 0$, *esiste* $\delta = \delta(\varepsilon) > 0$ *tale che*

$$|f(x) - f(x_0)| < \varepsilon\,, \ \forall x \in X \text{ con } |x - x_0| < \delta$$

iii) $\lim_{x \to x_0} f(x) = f(x_0)$

iv) per ogni successione $\{x_n\}$ *a valori in* X *e convergente a* x_0 *si abbia:*

$$\lim_{n \to +\infty} f(x_n) = f(\lim_{n \to +\infty} x_n) = f(x_0)\,.$$

Se si confronta la i) con la definizione 3.2.2. di limite di $f(x)$ per $x \to x_0$ si osserva che l'unica differenza è che ora la condizione $f(x) \in U$ deve essere verificata $\forall x \in V \cap X$, senza più escludere il punto x_0.

L'equivalenza di $i) \dots iv)$ è evidente (per la iv) ricordare il teorema 4.3.6). Le definizioni precedenti rendono rigorosa la definizione intuitiva: *una funzione è continua in* x_0 *se una piccola variazione del punto* x_0 *produce una piccola variazione dell'immagine* $f(x_0)$.

Se $f : X \to \mathbb{R}$ è continua in tutti i punti di un insieme $A \subseteq X$, diremo che è continua in A (o su A); diremo semplicemente che è continua se è continua in tutti i punti del suo insieme di definizione (ma vedi anche la seguente osservazione 1.1).

Esempi

Sono continue le funzioni seguenti:

1.1. Le funzioni costanti (ovvio!)

1.2. Le potenze: $x \mapsto x^\alpha$ $(x > 0, \alpha \in \mathbb{R})$; gli esponenziali: $x \mapsto a^x$ $(x \in \mathbb{R}$ e $a > 0)$; i logaritmi: $x \mapsto \log_a x$ $(x > 0, a > 0, a \neq 1)$: cfr. gli esempi 4.2.23, 4.2.24, 4.2.25.

1.3. $x \mapsto \sin x$. Infatti, $\forall x_0 \in \mathbb{R}$, abbiamo, posto $x = x_0 + h$:

$$\sin x - \sin x_0 = \sin(x_0 + h) - \sin x_0 = 2\sin(h/2)\cos(x_0 + h/2)\,;$$

poiché $|\cos(x_0 + h/2)| \leq 1$ e $\sin(h/2) \to 0$ per $h \to 0$, segue che $\sin x \to \sin x_0$ per $x \to x_0$.

1.4. $x \mapsto \cos x$. Procedere come nell'esempio precedente.

Come abbiamo parlato di limiti destro, sinistro, così possiamo parlare di continuità da destra, da sinistra. Più precisamente, diremo che

f *è continua da destra in* x_0 *se:* $\lim_{x \to x_{0+}} f(x) = f(x_0)$

f *è continua da sinistra in* x_0 *se:* $\lim_{x \to x_{0-}} f(x) = f(x_0)\,.$

***Esempio* 1.5** - La funzione $x \mapsto (x)$ (mantissa di x) è continua in ogni punto $x \neq 0, \pm 1, \pm 2, \cdots$ ed è continua da destra su tutto $\mathbb{R}$.

Sia $f : X \to \mathbb{R}$. Se x_0 è un punto interno di X, evidentemente *f è continua in x_0 se e solo se è ivi continua da destra e da sinistra.*

Sia $X = [a,b]$. Dire che f è continua in a (in b) significa dire che f è continua da destra in a (da sinistra in b).

Esaminiamo ora come si comporta la proprietà di continuità rispetto alle operazioni algebriche e ad altre operazioni eseguibili sulle funzioni.

Siano f e g due funzioni: $X \to \mathbb{R}$, continue. È immediato verificare che ogni combinazione lineare $c_1 f(x) + c_2 g(x)$ risulta pure continua su X; questa proprietà elementare è importante perché permette di considerare la classe delle funzioni continue come uno spazio vettoriale lineare: i "punti" dello spazio sono le singole funzioni, lo $\mathbf{0}$ è la funzione identicamente nulla su X, l'opposto di f è la funzione $-f$. Tale spazio viene solitamente indicato con $C(X)$ o $C^0(X)$; avremo occasione di tornare in maniera più approfondita su questo argomento.

È anche facile mostrare che (lo studente lo faccia per esercizio) se $f, g : X \to \mathbb{R}$ sono continue, allora risultano continue anche:

i) $f \cdot g$

ii) f/g (dove è definita, cioè dove $g(x) \neq 0$)

iii) $|f|$, f_+, f_-

iv) $\mathrm{Max}\{f,g\}$, cioè la funzione $x \mapsto \max\{f(x), g(x)\}$,
$\mathrm{Min}\{f,g\}$, cioè la funzione $x \mapsto \min\{f(x), g(x)\}$.

Siano ora $f : A \to B$ e $g : B \to \mathbb{R}$ due funzioni e consideriamo la funzione composta $g \circ f : A \to \mathbb{R}$.

■ **Teorema 1.1** - *Se f è continua in $x_0 \in A$ e g è continua in $f(x_0) \in B$, allora $g \circ f$ è continua in x_0.*

Dimostrazione - Essendo g continua in $f(x_0)$, per ogni intorno U di $g(f(x_0))$ esiste un intorno W di $f(x_0)$ tale che $g(W \cap B) \subset U$; per la continuità di f in x_0, in corrispondenza di questo W esiste un intorno V di x_0 tale che $f(V \cap A) \subset W \cap B$; e perciò $g(f(V \cap A)) \subset U$, il che prova che $g \circ f$ è continua in x_0. □

Possiamo domandarci se, come per la composizione, così la continuità si conserva anche con l'operazione di inversione di funzioni. Ma si vede subito che ciò non è vero in generale, come mostra il seguente esempio di funzione continua e invertibile, con inversa non continua.

***Esempio* 1.6** - Sia $X = [0,1] \cup (2,3]$ ed $f : X \to \mathbb{R}$ la funzione indicata in figura 5.1a; f è continua in X ma $f^{-1} : [0,2] \to X$ non è continua in $x = 1$.

Il problema della continuità della funzione inversa sarà affrontato, da un punto di vista più generale, nella sez. 2. Risulterà allora che, se f è continua e invertibile, la continuità della funzione inversa è garantita in almeno due casi: quando $X = \mathrm{dom}(f)$ è un intervallo (limitato o no) o quando X è un compatto. C'è anche il

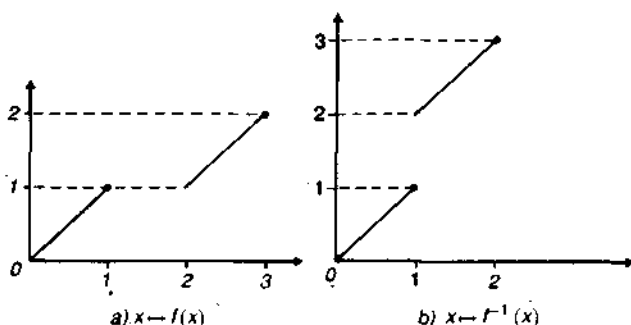



Fig. 5.1

caso estremamente importante di X aperto (teorema di "invarianza del dominio"); ma la dimostrazione di questo risultato comporta strumenti matematici che non possono essere presentati in questo libro.

Infine, il comportamento di una funzione continua rispetto all'operazione di limite è bene espresso dalla definizione stessa (iv): *le funzioni continue sono quelle per cui l'operazione di limite commuta con f.*

1.2 Punti di discontinuità

Sia $f : X \to \mathbb{R}$ ed esaminiamo i punti di X in cui la funzione non è continua; cerchiamo di capire quale può essere il comportamento della funzione in tali punti.

***Esempio* 1.7** - Sia $f(x) = \begin{cases} \sin x/x & \text{per } x \neq 0 \\ 0 & \text{per } x = 0 \end{cases}$; $f(x)$ è continua in ogni punto $x_0 \neq 0$ (poiché è il quoziente di due funzioni continue), ma non è continua in 0, poiché, per $x \to 0, f(x) \to 1 \neq f(0) = 0$. È chiaro che si tratta di una discontinuità fittizia; basterà infatti modificare la definizione della funzione nel punto 0, ponendo $f(x) = 1$ per $x = 0$, e la funzione sarà continua in quel punto. Parliamo in questo caso di *discontinuità eliminabile* (fig. 5.2a).

In generale diremo che x_0 (punto interno di $X = \text{dom}(f)$) è un punto di *discontinuità eliminabile* di $f(x)$ se esistono finiti i limiti destro e sinistro di f per $x \to x_0$ e:

$$\lim_{x \to x_{0+}} f(x) = \lim_{x \to x_{0-}} f(x) \neq f(x_0) .$$

Se $X = [a, b]$ e $x_0 = a$ si dirà che in a c'è una discontinuità eliminabile se $\lim\limits_{x \to a_+} f(x)$ esiste finito ed è diverso da $f(a)$ (analoga definizione per b).

***Esempio* 1.8** - La funzione $f(x) = \operatorname{sgn} x$ è discontinua in 0; tuttavia esistono finiti i limiti destro e sinistro (per $x \to 0$), ma sono diversi; in questo caso diremo che 0 è un punto di *salto* per $f(x)$ (fig. 5.2b).

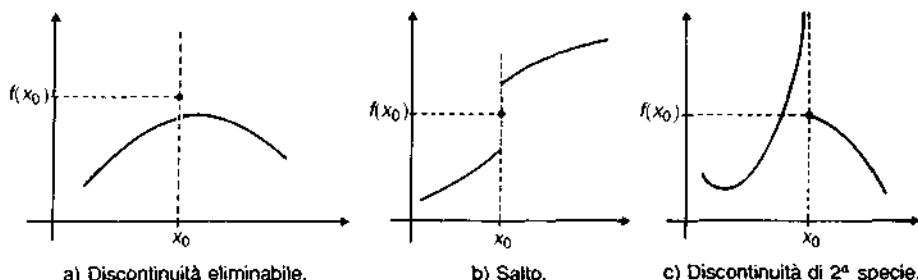



Fig. 5.2

In generale chiameremo *salto* di $f(x)$ per $x = x_0$ (x_0 punto interno di $X = \operatorname{dom}(f)$) la quantità

$$S(x_0) = \lim_{x \to x_{0+}} f(x) - \lim_{x \to x_{0-}} f(x) .$$

I punti con salto non nullo vengono anche detti punti di *discontinuità di 1ª specie*.

Un esempio interessante di funzioni che o sono continue o presentano solo discontinuità del tipo suaccennato è costituito dalle funzioni monotone in un intervallo chiuso $[a, b]$.

■ **Teorema 1.2** - *Una funzione $f : [a, b] \to \mathbb{R}$ monotona può avere al più una infinità numerabile di punti di discontinuità, e questi possono solo essere di 1ª specie (o eliminabili, se sono in a o in b).*

Dimostrazione - Sia, per esempio, f crescente; allora, per il teorema 4.2.10, sappiamo che, per ogni $x_0 \in (a, b)$, esistono finiti i limiti destro e sinistro e risulta

$$\lim_{x \to x_{0-}} f(x) \leq f(x_0) \leq \lim_{x \to x_{0+}} f(x)$$

cosicché o f è continua in x_0 (se i due limiti sono uguali) o presenta un salto in x_0. Analogamente si ha che: $\lim_{x \to a_+} f(x) \geq f(a)$ e $\lim_{x \to b_-} f(x) \leq f(b)$ cosicché questi punti o sono di continuità o di discontinuità eliminabile.

Poniamo ora $S = f(b) - f(a)$; possiamo osservare che in $[a, b]$ non potranno esservi più di $[S]$ (parte intera di S) punti con salto maggiore di 1, non potrannno esservi più di $2[S]$ punti con salto compreso tra $1/2$ e 1, non più di $4[S]$ punti con salto compreso tra $1/4$ e $1/2$ e così via; in questo modo si possono numerare (cioè ordinare in successione) tutti i punti di discontinuità di f, che risultano perciò al più un'infinità numerabile. □

***Esempio* 1.9** - Sia $f(x) = \begin{cases} e^{1/x} & \text{per } x \neq 0 \\ 0 & \text{per } x = 0 \end{cases}$; allora $\lim_{x \to 0_-} f(x) = 0$ e $\lim_{x \to 0_+} f(x) = +\infty$; perciò i limiti destro e sinistro esistono, ma uno di essi è infinito.

Esempio **1.10** - Sia $f(x) = \begin{cases} \sin 1/x & \text{per } x \neq 0 \\ 0 & \text{per } x = 0 \end{cases}$; in tal caso i limiti di $f(x)$, per $x \to 0_\pm$, non esistono.

In generale, quando uno (almeno) dei due limiti destro e sinistro è infinito, oppure non esiste, diremo che x_0 è un ***punto di discontinuità di 2^a^ specie*** (figura 5.2c).

Possiamo riassumere i vari tipi di discontinuità osservando che $\lim_{x \to x_0} f(x) =$ $= f(x_0)$ equivale a richiedere che i due limiti $\lim_{x \to x_0+} f(x)$ e $\lim_{x \to x_0-} f(x)$:

1) esistono finiti, 2) sono uguali e 3) sono uguali a $f(x_0)$.

Se 1) e 2) sono veri e 3) è falso, x_0è un punto di discontinuità eliminabile per f.
Se 1) è vero ma 2) (e quindi anche 3)) è falso, x_0 è un punto di salto o di discontinuità di 1^a specie.
Se 1) è falso, x_0 è di discontinuità di 2^a specie.

Osservazione 1.1 - Per poter dire che una funzione $f : X \to \mathbb{R}$ è continua oppure no in x_0 si richiede che f sia definita in x_0, cioè $x_0 \in X$. Così, a rigore, è bene evitare espressioni del tipo: la funzione $f(x) = 1/x$ (ovviamente definita per $x \neq 0$) ha in $x_0 = 0$ un punto di discontinuità (di 2^a specie), oppure: la funzione $f(x) = \sin x/x$ (anch'essa definita per $x \neq 0$) è continua in $x_0 = 0$. L'uso di tali locuzioni, che pure è frequente, è pertanto ammissibile solo se si sottintende che le funzioni in questione siano state tacitamente definite in x_0 in modo opportuno.
Così si potrà dire che $f(x) = 1/x$ è continua (dove è definita, cioè per $x \neq 0$) e anche che 0 è un punto di discontinuità (di 2^a specie) per $f(x)$. Analogamente diremo che $\log x$ è continua (per $x > 0$) e che 0 è punto di discontinuità (di 2^a specie) per $\log x$.
Se invece i due limiti (destro e sinistro) esistono finiti e sono uguali, diremo che $f(x)$ è continua in x_0, anche se non è ivi definita. Così diremo che la funzione $f(x) = \exp(-1/x^2)$ per $x \neq 0$ (e non definita in $x = 0$) è continua su tutto $\mathbb{R}$, anche in 0, poiché $\lim_{x \to 0_+} f(x) = \lim_{x \to 0_-} f(x) = 0$.

Osservazione 1.2 - Anche se, nella maggior parte delle applicazioni, i punti di discontinuità sono un po' un'eccezione, mentre la continuità è la regola, esistono funzioni che non sono continue in alcun punto; per esempio l'insieme dei punti di discontinuità della funzione di Dirichlet: $f(x) = 1$ se $x \in \mathbb{Q}$, $f(x) = 0$ se $x \in \mathbb{R} \setminus \mathbb{Q}$, coincide con il suo insieme di definizione, cioè $\mathbb{R}$. Di che specie sono queste discontinuità?

1.3 Proprietà fondamentali delle funzioni continue su un intervallo

Passiamo ora in rassegna le proprietà fondamentali delle funzioni definite su un intervallo e ivi continue; non dimostreremo in questo paragrafo tutti i risultati

che enunceremo, poiché alcuni di questi sono dei semplici corollari di teoremi dimostrati nella sez. 2.

Cominciamo con le proprietà in qualche modo connesse con l'ordinamento di ℝ.

La prima osservazione (e questa vale per le funzioni continue su un insieme X qualsiasi) è che, essendo, per una funzione continua in x_0, $f(x_0) = \lim_{x \to x_0} f(x)$, il teorema della permanenza del segno per i limiti (teorema 4.2.3) ha come immediata conseguenza il seguente

■ **Teorema 1.3** - (della permanenza del segno)

Sia $f : X \to \mathbb{R}$ una funzione continua in $x_0 \in X$. Se $f(x_0) > 0$, allora esiste un intorno U di x_0 tale che $f(x) > 0$ per ogni $x \in U \cap X$.

Considerando la funzione $f(x) - l$ ($l \in \mathbb{R}$) segue immediatamente da quanto sopra che, se $f(x_0) > l$, allora si mantiene $f(x) > l$ in tutto un intorno di x_0. Molto importante è il seguente

■ **Teorema 1.4** - (degli zeri)

Sia $f : [a,b] \to \mathbb{R}$ continua e risulti $f(a)f(b) < 0$. Allora esiste (almeno) un punto ξ compreso tra a e b tale che $f(\xi) = 0$.

Dimostrazione - Questo teorema potrà dedursi come corollario dal teorema 2.8. Tuttavia ne diamo qui una dimostrazione indipendente, di tipo costruttivo: si prova l'esistenza dello zero fornendo un algoritmo che lo determina; utilizzeremo il procedimento di dicotomia.

Sia, per esempio, $f(a) > 0$, $f(b) < 0$. Detto c il punto medio di $[a,b]$, $c = (a+b)/2$, consideriamo $f(c)$; se $f(c) = 0$, il teorema è dimostrato; se $f(c) > 0$ consideriamo l'intervallo $[c,b]$; se $f(c) < 0$ consideriamo l'intervallo $[a,c]$.

Detto $[a_1,b_1]$ quello dei due intervalli che abbiamo così ottenuto, osserviamo che risulta: $f(a_1) > 0$, $f(b_1) < 0$. Ripetiamo su $[a_1,b_1]$ la stessa operazione fatta su $[a,b]$ e così via.

Così procedendo, può accadere che uno dei punti medi degli intervalli che così si ottengono sia uno zero di $f(x)$ e allora il teorema è dimostrato; oppure ciò non avviene mai, e allora si ottengono due successioni $\{a_n\}$ e $\{b_n\}$ tali che

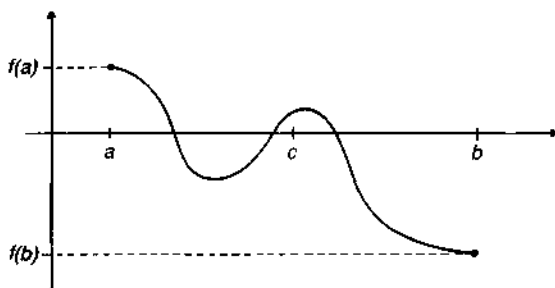



Fig. 5.3

$$a \le a_1 \le a_2 \le \cdots \le a_n \le \cdots \le b_n \le \cdots \le b_2 \le b_1 \le b$$

$$f(a_n) > 0 \quad , \quad f(b_n) < 0 \quad \text{per ogni } n$$

$$b_n - a_n = \frac{b-a}{2^n} \ . \tag{1.1}$$

La successione $\{a_n\}$ ammette limite finito (poiché è monotona crescente e limitata superiormente da b); la successione $\{b_n\}$ pure ammette limite finito (è monotona decrescente e limitata inferiormente da a) e i due limiti devono essere uguali per la 1.1. Sia $\xi = \lim_{n\to+\infty} a_n = \lim_{n\to+\infty} b_n$. Essendo f continua in ξ risulta allora simultaneamente

$$f(\xi) = \lim_{n\to+\infty} f(a_n)$$
$$f(\xi) = \lim_{n\to+\infty} f(b_n) \ .$$

Ma $f(a_n) > 0$ e perciò $f(\xi) \ge 0$ (ricordare l'osservazione 4.2.3) $f(b_n) < 0$ e perciò $f(\xi) \le 0$ da cui segue $f(\xi) = 0$. □

Si osservi che, arrestando il procedimento dopo n passi, si può assumere a_n (b_n) come valore approssimato di ξ, commettendo un errore per difetto (eccesso) non superiore a $(b-a)/2^n$.

Si osservi che, se in $[a, b]$ vi sono più zeri della funzione f, il procedimento illustrato non ci dice quale di essi viene determinato.

***Esempio* 1.11** - Si voglia determinare, con un calcolatore tascabile, una radice dell'equazione: $x = 2^{-x}$.

Posta l'equazione nella forma:

$$f(x) = 2^{-x} - x = 0$$

si vede che essendo $f(0) = 1$ e $f(1) = -1/2$, nell'intervallo $[0,1]$ c'è una radice (è anche facile convincersi che è unica!). Applichiamo il procedimento sopra descritto

$$\begin{array}{lll}
f(1/2) = 0.2071 > 0 & \text{considero } [1/2, 1] \\
f(3/4) = -0.1554 < 0 & \text{considero } [1/2, 3/4] \\
f(5/8) = 0.2920 > 0 & \text{considero } [5/8, 3/4] \\
f(11/16) = -0.0666 < 0 & \text{considero } [5/8, 11/16] \\
f(21/32) = -0.0217 < 0 & \text{considero } [5/8, 21/32] \\
f(41/64) = 0.0008 > 0 &
\end{array}$$

La radice cercata, ξ, si trova nell'intervallo $[41/64, 21/32]$, e cioè abbiamo $0.640625 < \xi < 0.656225$ ovvero $\xi = 0.6...$ con la seconda cifra incerta tra 4 e 5, dopo 6 iterazioni del procedimento. Il metodo è dunque poco efficiente.

Assumendo l'uno o l'altro dei due valori approssimati (per difetto o per eccesso) si commette un errore non superiore a $1/2^6 = 0.015625$.

Il precedente teorema viene a cadere se f non è continua oppure se X non è un intervallo: lo studente trovi dei controesempi.

Corollario 1.5 - *Una funzione continua in un intervallo* (a, b) *assume tutti i valori compresi tra* $\inf f$ *e* $\sup f$.

Dimostrazione - Sia t un numero compreso tra $\inf f$ e $\sup f$. Allora esisteranno due punti in (a, b), x_1 e x_2, tali che:

$$\inf f \leq f(x_1) < t < f(x_2) \leq \sup f$$

La funzione $g(x) = f(x) - t$ è continua in $[x_1, x_2]$ ed inoltre $g(x_1) < 0$, $g(x_2) > 0$. Esiste dunque ξ tale che $g(\xi) = 0$ e cioè esiste ξ tale che: $f(\xi) = t$. □

***Esempio* 1.12** - La funzione $f(x) = \operatorname{tg} x$ da $(-\pi/2, +\pi/2)$ in $\mathbb{R}$ è continua (poiché è il quoziente di due funzioni continue con denominatore diverso da zero) e inoltre risulta: $\inf f = -\infty$, $\sup f = +\infty$. Allora l'immagine di f è *tutto* $\mathbb{R}$.

Proprietà ancora più significative si possono ottenere se consideriamo le *funzioni continue su un intervallo chiuso e limitato*; esse vengono qui sotto enunciate in una Proposizione, che è un semplice Corollario del Teorema 2.4 della sez. 2 (vedi Corollario 2.5).

Proposizione 1.6 - Sia $f : [a, b] \to \mathbb{R}$ *continua. Allora:*

1. f è limitata in $[a, b]$;

2. f ammette massimo e minimo;

3. f assume (almeno una volta) qualunque valore compreso tra il minimo e il massimo (proprietà di Darboux o dei valori intermedi).

Si osservi che la proprietà 3. discende dal Corollario 1.5 e dalla proprietà 2.

È facile convincersi che le proprietà 1., 2., 3., vengono generalmente (non necessariamente) a cadere se cade qualche ipotesi del teorema, come mostrano gli esempi seguenti.

Esempi

1.13. $(-\pi/2, +\pi/2) \ni x \mapsto \operatorname{tg} x$: funzione continua in un intervallo limitato, non chiuso. La funzione non è limitata, non ammette massimo né minimo.

1.14. $[0, +\infty) \ni x \mapsto x^2$: funzione continua in un intervallo chiuso (in $\mathbb{R}$) ma non limitato. La funzione non è limitata, ammette minimo, non ammette massimo.

1.15. $[-1, +1] \ni x \mapsto \operatorname{sgn} x$: funzione non continua in un intervallo chiuso e limitato. La funzione non possiede la proprietà di Darboux.

1.16. $[0, 1) \ni x \mapsto f(x) = \begin{cases} \sin 1/x & \text{per } x \neq 0 \\ 0 & \text{per } x = 0 \end{cases}$: funzione non continua in un intervallo non chiuso. La funzione possiede tutte le proprietà elencate nella Proposizione 1.6.

1.4 La continuità uniforme

Consideriamo, nei due esempi seguenti, due funzioni, entrambe continue, che presentano però una importante differenza.

***Esempio* 1.17** - Sia $f(x) = x^2$. Questa funzione è continua in $(0,1)$; infatti, fissato $\varepsilon > 0$, per ogni $x_0 \in (0,1)$, prendiamo $\delta = \varepsilon/2$. Allora risulta, se $|x-x_0| < \delta$,

$$|f(x) - f(x_0)| = |x^2 - x_0^2| = |x - x_0|\,|x + x_0| \leq \frac{\varepsilon}{2}(|x| + |x_0|) \leq \varepsilon \; .$$

Notiamo che abbiamo trovato un δ adatto per tutti gli x_0 di $(0,1) : \delta = \delta(\varepsilon)$, indipendente da x_0.

***Esempio* 1.18** - Sia $g(x) = 1/x$. Anch'essa è continua in $(0,1)$; infatti, fissiamo $\varepsilon > 0$ e $x_0 \in (0,1)$. Se vogliamo avere

$$|g(x) - g(x_0)| = \left|\frac{1}{x} - \frac{1}{x_0}\right| \leq \varepsilon \tag{1.2}$$

dovrà essere, risolvendo l'ultima disequazione,

$$\frac{x_0}{1 + \varepsilon x_0} \leq x \leq \min\left\{1, \frac{x_0}{1 - \varepsilon x_0}\right\} .$$

Ciò significa che la scelta migliore per δ (cioè il δ più grande per cui, se $|x-x_0| < \delta$, vale la (1.2)) sarà:

$$\delta = \delta(\varepsilon, x_0) = \frac{\varepsilon x_0^2}{1 + \varepsilon x_0} .$$

Si noti che, quando x_0 si avvicina a 0, anche δ si avvicina a 0.

La differenza tra questo caso e il precedente sta nel fatto che ora non possiamo più scegliere un δ appropriato per *tutti* gli x_0 di $(0,1)$, pur essendo naturalmente possibile scegliere un δ per *ogni singolo* x_0; abbiamo cioè $\delta = \delta(\varepsilon, x_0)$ ma non $\delta = \delta(\varepsilon)$, indipendente da x_0!

La differenza tra le due situazioni sopra illustrate ha importanti conseguenze, che potranno però essere pienamente apprezzate dall'allievo solo più avanti. Ora introduciamo un concetto, la continuità uniforme, che ci consente di precisare e generalizzare quanto sopra illustrato.

Definizione 1.2 - *Sia $X \subseteq \mathbb{R}$ e $f : X \to \mathbb{R}$. Si dice che f è uniformemente continua in X se, $\forall \varepsilon > 0$, $\exists \delta > 0$ ($\delta = \delta(\varepsilon)$) tale che: per ogni coppia di punti $x, x_0 \in X$, con*

$$|x - x_0| < \delta \quad \textit{risulti} \quad |f(x) - f(x_0)| < \varepsilon \; .$$

Confrontiamo questa definizione con la definizione di funzione continua in X: $\forall \varepsilon > 0$, $\forall x_0 \in X$, $\exists \delta > 0$ $(\delta = \delta(\varepsilon, x_0))$ *tale che: per ogni* $x \in X$, *con*

$$|x - x_0| < \delta \quad \textit{risulti} \quad |f(x) - f(x_0)| < \varepsilon \; .$$

È evidente dunque che, se f è uniformemente continua in X, è ivi anche continua; ma il viceversa non è vero; i due casi illustrati sopra mostrano che la funzione x^2 è uniformemente continua in $(0,1)$, la funzione $1/x$ non lo è (pur essendo continua).

Esempi

1.19. La funzione $f(x) = x^2$ non è uniformemente continua in $[0, +\infty)$. Infatti, posto $x = x_0 + h$, risulta $f(x) - f(x_0) = (x_0 + h)^2 - x_0^2 = 2hx_0 + h^2 > 2hx_0$.
Fissato comunque h, si può sempre trovare un x_0 tale che $2hx_0 > 1$ (basta prendere $x_0 > 1/2h$). Ciò mostra che, fissato ε, è impossibile trovare un δ dipendente solo da ε (e non da x_0) per cui risulti $|f(x) - f(x_0)| < \varepsilon$ quando $|x - x_0| < \delta$.

1.20. La funzione e^{-x} è uniformemente continua in $[0, +\infty)$. Infatti, posto $x = x_0 + h$, risulta

$$|e^{-x} - e^{-x_0}| = e^{-x_0}|e^{-h} - 1| \leq |e^{-h} - 1| < \varepsilon$$

purché $|h| = |x - x_0| < \min\{\log(1 + \varepsilon), -\log(1 - \varepsilon)\}$.

La continuità uniforme può essere illustrata in termini di oscillazione di una funzione.
Se $f : X \to \mathbb{R}$ ed $E \subseteq X$ definiamo **oscillazione** di f in E la quantità

$$\omega_E(f) = \sup_E f - \inf_E f \; .$$

La definizione 1.2 equivale a richiedere che, fissato $\varepsilon > 0$, esiste $\delta = \delta(\varepsilon)$ tale che per *ogni* $E \subseteq X$ con diametro minore di δ, $\omega_E(f) < \varepsilon$.
Così, ad esempio, se $X = (a, b)$ ed f è uniformemente continua in (a, b), fissato ε è possibile suddividere l'intervallo (a, b) in un numero finito di sottointervalli di ampiezza δ in modo che su ciascuno di essi l'oscillazione di f è più piccola di ε.
Le proprietà elencate nella Proposizione 1.6 sono legate alla continuità uniforme dal seguente teorema (di Cantor–Heine), che sarà dimostrato nella sez. 2.

Proposizione 1.7 - *Una funzione* $f : [a, b] \to \mathbb{R}$ *continua (su un intervallo chiuso e limitato) è uniformemente continua.*

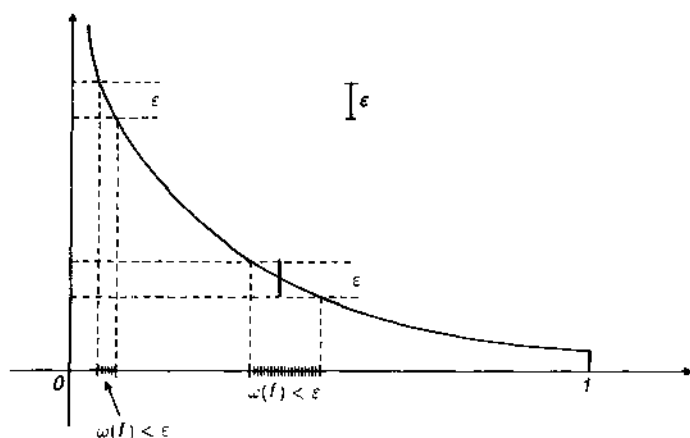



Fig. 5.4 Funzione non uniformemente continua in $(0,1)$: per mantenere l'oscillazione minore di ε vicino a 0 occorre scegliere insiemi di diametro sempre più piccolo.

Naturalmente, non è necessario che il dominio di definizione di una funzione continua sia chiuso e limitato perché essa sia uniformemente continua, come mostra l'esempio 1.20. Si noti anche che una funzione costante è uniformemente continua su ogni insieme X.

Condizioni sufficienti per la continuità uniforme sono illustrate negli esercizi 1.6, 1.7, 1.8.

Esercizi

1. Una funzione $f : X \to \mathbb{R}$ si dice *semicontinua superiormente* in un punto $x_0 \in X$ se

$$\overline{\lim_{x\to x_0}}\, f(x) \le f(x_0)\ ;$$

si dice *semicontinua inferiormente* in x_0 se

$$\underline{\lim_{x\to x_0}}\, f(x) \ge f(x_0)\ .$$

Mostrare che una funzione è continua in un punto se e solo se è semicontinua inferiormente e superiormente in quel punto.

2. Verificare che la funzione:

a) $\operatorname{sgn} x$ non è semicontinua né inferiormente né superiormente in 0

b) $[x]$ è semicontinua superiormente in ogni punto di $\mathbb{R}$

c) (x) è semicontinua inferiormente in ogni punto di $\mathbb{R}$.

3. Data la funzione: $H(x) = 0$ se $x < 0$, $H(x) = 1$ se $x > 0$ (gradino di Heaviside), definire H anche in 0 in modo che risulti semicontinua inferiormente (superiormente).

4. Trovare, per ciascuna delle seguenti equazioni, aiutandosi con dei grafici, un intervallo contenente una soluzione; quindi, eventualmente con l'uso di un calcolatore tascabile, e usando il procedimento di dicotomia, determinare una soluzione con buona approssimazione:

a) $\operatorname{tg} x = e^x$

b) $\log |x| = \cos x$

c) $\dfrac{\sin x}{x} = \sqrt{|x|}$.

5. Una funzione $f : X \to \mathbb{R}$ si dice *hölderiana* di ordine $\alpha : 0 < \alpha \leq 1$ in $A \subseteq X$ se esiste una costante positiva M tale che:

$$|f(x) - f(y)| \leq M|x - y|^\alpha \qquad \forall\, x, y \in A\ .$$

Se f è hölderiana di ordine 1 viene detta *lipschitziana*. Evidentemente una funzione hölderiana è continua; ma non è vero il viceversa. L'ordine di hölderianità α è una specie di misura quantitativa della continuità.

Dire se le seguenti funzioni sono hölderiane, e calcolarne, in caso positivo, l'ordine:

a) $\sqrt{x}$ in $[0, 1]$

b) $\sin(\sqrt[3]{x})$ in $\mathbb{R}$

c) $|x|^\alpha$ in $[-1, 1], \alpha \geq 0$

d) $1/\log x$ in $(0, 1)$.

6. Dimostrare che una funzione hölderiana in A è uniformemente continua in A.

7. Dimostrare che se $f : [a, +\infty) \to \mathbb{R}$ è continua e ammette asintoto orizzontale, allora f è uniformemente continua.

8. Dimostrare che se $f : [a, +\infty) \to \mathbb{R}$ è continua e ammette asintoto obliquo, allora f è uniformemente continua.

9. Dire se le funzioni

a) $\sqrt{x}$

b) $x \log x$

c) e^x

sono uniformemente continue in $[0, +\infty)$.

2. FUNZIONI CONTINUE DA $\mathbb{R}^n$ IN $\mathbb{R}^m$

2.1 Una caratterizzazione delle funzioni continue

Ora che lo studente si è familiarizzato con il concetto di funzione continua nel caso particolare di funzioni da $\mathbb{R}$ in $\mathbb{R}$, possiamo agevolmente affrontare lo studio della continuità da un punto di vista più generale.

Sia perciò $X \subseteq \mathbb{R}^n$ e $\mathbf{f} : X \to \mathbb{R}^m$. La definizione di continuità in un punto $\mathbf{x}^0 \in X \cap X'$ e in un insieme $A \subseteq X$ si ripete pari pari come in 1.1. Così pure si estendono le affermazioni circa la continuità della somma e del prodotto di funzioni continue e la continuità della funzione composta.

Le proprietà generali delle funzioni continue possono essere divise in due gruppi, secondoché la funzione sia definita su un compatto o su un connesso.

L'importanza degli insiemi *connessi* e degli insiemi *compatti* in rapporto alle funzioni continue deriva dal fatto che, come vedremo, *le proprietà di compattezza e connessione vengono preservate da una trasformazione continua*, cioè una funzione continua (da $\mathbb{R}^n$ in $\mathbb{R}^m$) manda compatti di $\mathbb{R}^n$ in compatti di $\mathbb{R}^m$, e lo stesso avviene per i connessi.

Si noti che, in generale, una funzione continua non manda un aperto in un aperto, né un chiuso in un chiuso, come prova l'esempio seguente:

$$f : x \mapsto \frac{1}{1+x^2} \ ;$$

f manda $\mathbb{R}$ (insieme aperto e chiuso) nell'intervallo $(0, 1]$ (né aperto, né chiuso).

A questo proposito, una caratterizzazione delle funzioni continue è possibile introducendo nel loro dominio la cosiddetta "**topologia relativa o indotta**".

Sia $X \subseteq \mathbb{R}^n$ e $\mathbf{x}^0 \in X$. Si dice *intorno* di $\mathbf{x}^0$ nella topologia indotta o relativa di X, o semplicemente, *intorno* di $\mathbf{x}_0$ in X, un insieme della forma $U \cap X$ dove U è un intorno di $\mathbf{x}_0$ in $\mathbb{R}^n$. È facile verificare che il sistema di insiemi così definito soddisfa le proprietà $\mathcal{U}_1, \cdots, \mathcal{U}_4$, caratteristiche degli intorni.

Si potrà così parlare di aperti, chiusi, punti di accumulazione etc., rispetto alla topologia indotta di X; diremo semplicemente aperti, chiusi, punti di accumulazione ... in X.

Ad esempio gli aperti in X sono della forma $A \cap X$ con A aperto in $\mathbb{R}^n$; analogamente per i chiusi.

Vale il seguente teorema.

▪ **Teorema 2.1 -** $\mathbf{f} : \mathbb{R}^n \supseteq X \to \mathbb{R}^m$ *è continua se e solo se la controimmagine* $\mathbf{f}^{-1}(A)$ *di ogni aperto non vuoto di* $\mathbb{R}^m$ *è un aperto in* X.

Dimostrazione - *i*) Sia $\mathbf{f}$ continua in X e sia $A \subseteq \mathbb{R}^m$ un aperto non vuoto. Mostriamo che $\mathbf{f}^{-1}(A)$ è un aperto in X. Se $\mathbf{f}^{-1}(A) = \emptyset$ la tesi è provata. Diversamente, sia $\mathbf{x}^0 \in \mathbf{f}^{-1}(A)$, cioè $\mathbf{f}(\mathbf{x}^0) \in A$; essendo A aperto, $\mathbf{f}(\mathbf{x}^0)$ è un punto interno di A e perciò esiste un intorno U di $\mathbf{f}(\mathbf{x}^0)$ contenuto in A. Ma, essendo $\mathbf{f}$ continua in $\mathbf{x}^0$, possiamo trovare un intorno W di $\mathbf{x}^0$ tale che $\mathbf{f}(\mathbf{x}) \in U$ se $\mathbf{x} \in W \cap X$, e perciò $\mathbf{f}(\mathbf{x}) \in A$, cioè $\mathbf{x} \in \mathbf{f}^{-1}(A)$; in conclusione, se $\mathbf{x}^0 \in \mathbf{f}^{-1}(A)$, esiste un suo intorno $W \cap X$, in X, tutto contenuto in $\mathbf{f}^{-1}(A)$ e perciò $\mathbf{f}^{-1}(A)$ è un aperto in X.

ii) Sia $\mathbf{f}^{-1}(A)$ aperto in X se A è aperto in $\mathbb{R}^m$. Sia $\mathbf{x}^0 \in X$ e mostriamo che $\mathbf{f}$ è continua in $\mathbf{x}^0$. Fissato un intorno A di $\mathbf{f}(\mathbf{x}^0)$, risulta $\mathbf{f}^{-1}(A)$ aperto in X; poiché $\mathbf{x}^0 \in \mathbf{f}^{-1}(A)$ possiamo trovare un intorno W di $\mathbf{x}^0$ contenuto in $\mathbf{f}^{-1}(A)$, cioè tale che $f(W \cap X) \subseteq A$, il che prova la continuità di $\mathbf{f}$. □

Corollario 2.2 - *$\mathbf{f} : \mathbb{R}^n \supseteq X \to \mathbb{R}^m$ è continua se e solo se la controimmagine $\mathbf{f}^{-1}(C)$ di ogni chiuso non vuoto di $\mathbb{R}^m$ è un chiuso in X.*

Per la dimostrazione, che lasciamo per esercizio, utilizzare la legge di dualità.

L'esempio più semplice di funzioni continue è dato dalle *funzioni lineari*.

Definizione 2.1 - *Una funzione $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ si dice additiva se:*

$$\mathbf{f}(\mathbf{x}+\mathbf{y}) = \mathbf{f}(\mathbf{x}) + \mathbf{f}(\mathbf{y}) \qquad \forall\, \mathbf{x},\mathbf{y} \in \mathbb{R}^n ; \tag{2.1}$$

si dice omogenea se:

$$\mathbf{f}(c\mathbf{x}) = c\mathbf{f}(\mathbf{x}) \qquad \forall\, \mathbf{x} \in \mathbb{R}^n,\ \forall\, c \in \mathbb{R} ; \tag{2.2}$$

si dice lineare se è additiva e omogenea.

Proposizione 2.3 - *Ogni funzione lineare è continua.*

Dimostrazione - Per l'additività, è sufficiente mostrare che $\mathbf{f}$ è continua in $\mathbf{0}$, poiché, per mostrare che è continua in $\mathbf{x}^0$ si scrive:

$$\mathbf{f}(\mathbf{x}^0 + \mathbf{h}) - \mathbf{f}(\mathbf{x}^0) = \mathbf{f}(\mathbf{h}) .$$

Posto allora $\mathbf{h} = ||\mathbf{h}||\mathbf{e}$ ($\mathbf{e}$ versore) abbiamo, per la (2.2)

$$\lim_{\mathbf{h}\to\mathbf{0}} \mathbf{f}(\mathbf{h}) = \lim_{\mathbf{h}\to\mathbf{0}} \mathbf{f}(||\mathbf{h}||\mathbf{e}) = \lim_{\mathbf{h}\to\mathbf{0}} ||\mathbf{h}||\mathbf{f}(\mathbf{e}) = \mathbf{0}$$

e poiché $\mathbf{f}(\mathbf{0}) = \mathbf{0}$ (basta porre $\mathbf{y} = \mathbf{0}$ nella (2.1)) $\mathbf{f}$ è continua in $\mathbf{0}$. □

Le uniche funzioni lineari da $\mathbb{R}$ in $\mathbb{R}$ sono della forma: $f(x) = ax$ ($a \in \mathbb{R}$), cioè sono le rette per l'origine; ciò discende dalla sola proprietà di omogeneità, ponendo $x = 1$ nella (2.2).

Le funzioni lineari da $\mathbb{R}^n$ in $\mathbb{R}$ sono i piani (iperpiani) passanti per l'origine, cioè della forma:

$$f(\mathbf{x}) = a_1x_1 + a_2x_2 + \cdots + a_nx_n = <\mathbf{a},\mathbf{x}>$$

con $\mathbf{a} \in \mathbb{R}^n$.

E infine le funzioni (o trasformazioni) lineari da $\mathbb{R}^n$ in $\mathbb{R}^m$ sono della forma:

$$\mathbf{f}(\mathbf{x}) = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{pmatrix}$$

cioè $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$, con $\mathbf{A}$ matrice $m \times n$.

Più precisamente, data una funzione $\mathbf{f}$, lineare da $\mathbb{R}^n$ in $\mathbb{R}^m$, e fissato il sistema di riferimento in $\mathbb{R}^n$ ed $\mathbb{R}^m$, esiste un'unica matrice $\mathbf{A}$, $m \times n$, tale che $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x}$.

Osservazione 2.1 - Per un abuso di linguaggio talvolta si chiamano lineari le funzioni:

$$\mathbb{R} \to \mathbb{R} : x \mapsto ax + b$$

$$\mathbb{R}^n \to \mathbb{R} : \mathbf{x} \mapsto a_1 x_1 + \cdots + a_n x_n + b$$

$$\mathbb{R}^n \to \mathbb{R}^m : \mathbf{x} \mapsto \mathbf{A}\mathbf{x} + \mathbf{b}$$

e si chiamano lineari omogenee quelle che qui sono chiamate lineari.

Lo studio delle funzioni lineari (o trasformazioni lineari) da $\mathbb{R}^n$ in $\mathbb{R}^m$ viene svolto nei corsi di Algebra lineare. Gli elementi base di questa teoria, e cioè i teoremi di Cramer e di Rouché-Capelli sulla iniettività e suriettività di tali trasformazioni, devono essere familiari allo studente.

***Esempio* 2.1** - La funzione $f : \mathbb{R}^2 \to \mathbb{R}$ così definita: $f(x_1, x_2) = (x_1^2 - x_2^2)/(x_1^2 + x_2^2)$ per $(x_1, x_2) \neq (0,0)$, $f(0,0) = 0$ non è continua in $(0,0)$; infatti non esiste il limite di $f(\mathbf{x})$ per $\mathbf{x} \to \mathbf{0}$.

***Esempio* 2.2** - La funzione $f : \mathbb{R}^n \to \mathbb{R}$ così definita:

$$f(\mathbf{x}) = \begin{cases} \exp\left\{-\dfrac{1}{1 - \|\mathbf{x}\|^2}\right\} & \text{se } \|\mathbf{x}\| < 1 \\ 0 & \text{se } \|\mathbf{x}\| \geq 1 \end{cases}$$

è continua in tutto $\mathbb{R}^n$. Infatti f è una funzione radiale (dipende cioè solo da $\|\mathbf{x}\|$) e quindi posto $\rho = \|\mathbf{x}\|$ abbiamo

$$f(\mathbf{x}) = f(\rho) = \begin{cases} \exp\{-\dfrac{1}{1-\rho^2}\} & \text{se } \rho < 1 \\ 0 & \text{se } \rho \geq 1 \end{cases}.$$

I soli punti di discontinuità possono essere quelli della ipersuperficie sferica di equazione $\rho = 1$. Ma se $\|\mathbf{x}^0\| = 1$, $\mathbf{x} \to \mathbf{x}^0$ implica $\rho \to 1$ e allora $f(\rho) \to 0 = f(\mathbf{x}^0)$; quindi f è continua anche nei punti di modulo unitario.

Osservazione 2.2 - Come abbiamo già osservato a proposito dei limiti, così la continuità di una funzione $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m$ equivale alla continuità delle sue componenti $(f_1, f_2, \cdots, f_m)$.

In particolare, la continuità delle funzioni di variabile complessa: $z \to f(z)$, dove $z = x + iy$, $f(z) = u(x,y) + iv(x,y)$, $(x, y \in \mathbb{R})$ equivale alla continuità della parte reale $u : \mathbb{R}^2 \to \mathbb{R}$ e della parte immaginaria $v : \mathbb{R}^2 \to \mathbb{R}$.

***Esempio* 2.3** - Le funzioni $f : \mathbb{C} \to \mathbb{C}$ così definite:

i) $z \mapsto \overline{z}$

ii) $z \to |z|$

iii) $z \to z^2$

sono continue su tutto $\mathbb{C}$.

2.2 Funzioni continue su un compatto

■ **Teorema 2.4** - *Sia $K \subset \mathbb{R}^n$ compatto; sia $\mathbf{f} : K \to \mathbb{R}^m$ continua; allora $\mathbf{f}(K)$ è compatto.*

Dimostrazione - Dimostriamo che $\mathbf{f}(K)$ è compatto per successioni e cioè che da ogni successione di elementi di $\mathbf{f}(K)$ si può estrarre una sottosuccessione convergente in $\mathbf{f}(K)$.

Sia $\{\mathbf{y}_k\}$ una successione in $\mathbf{f}(K)$. Sia $\{\mathbf{x}_k\}$ una successione in K tale che: $\mathbf{f}(\mathbf{x}_k) = \mathbf{y}_k$. Poiché K è compatto, dalla $\{\mathbf{x}_k\}$ si può estrarre una sottosuccessione $\{\mathbf{x}_{k_l}\}$ convergente a un certo $\mathbf{x} \in K$. Allora risulta che la successione $\{\mathbf{y}_{k_l}\}$, con $\mathbf{y}_{k_l} = \mathbf{f}(\mathbf{x}_{k_l})$ è una sottosuccessione di $\{\mathbf{y}_k\}$ convergente a $\mathbf{f}(\mathbf{x})$ per la continuità di $\mathbf{f}$. Ciò prova che $\mathbf{f}(K)$ è compatto. □

Corollario 2.5 - (Teorema di Weierstrass)

Sia $K \in \mathbb{R}^n$ compatto; sia $\mathbf{f} : K \to \mathbb{R}$ continua; allora f ammette massimo e minimo.

Dimostrazione - $\mathbf{f}(K)$ è compatto, cioè chiuso e limitato. Essendo limitato, risultano finiti l'estremo superiore e inferiore di $\mathbf{f}$:

$$\inf_K \mathbf{f} = m \qquad \sup_K \mathbf{f} = M \ .$$

Essendo chiuso, esiste almeno un punto in K in cui f assume il valore m e un punto in cui assume il valore M (ricorda l'Osservazione 3.2.3). □

Risulta così provata la Proposizione1.6.

■ **Teorema 2.6** - *Sia $K \subset \mathbb{R}^n$ compatto; sia $\mathbf{f} : K \to \mathbb{R}^m$ iniettiva e continua. Allora $\mathbf{f}^{-1} : \mathbf{f}(K) \to K$ è continua.*

Dimostrazione - Sia $\mathbf{y} \in \mathbf{f}(K)$ e $\{\mathbf{y}_k\}$ una successione a valori in $\mathbf{f}(K)$ convergente a $\mathbf{y}$; dimostriamo che la successione $\mathbf{x}_k = \mathbf{f}^{-1}(\mathbf{y}_k)$ converge a $\mathbf{x} = \mathbf{f}^{-1}(\mathbf{y})$. Sia $\mathbf{z}$ un valore limite di $\{\mathbf{x}_k\}$; essendo $\{\mathbf{x}_k\} \subset K$, con K compatto, deve essere $\mathbf{z} \in K$. Sia $\mathbf{x}_{k_l}$ una sottosuccessione estratta da $\{\mathbf{x}_k\}$ e convergente a $\mathbf{z}$. Dalla continuità di f segue allora che $\mathbf{f}(\mathbf{x}_{k_l}) = \mathbf{y}_{k_l} \to \mathbf{f}(\mathbf{z})$. Poiché $\mathbf{y}_k \to \mathbf{y}$, anche $\mathbf{y}_{k_l} \to \mathbf{y} = \mathbf{f}(\mathbf{x})$. Per l'unicità del limite deve essere $\mathbf{f}(\mathbf{z}) = \mathbf{f}(\mathbf{x})$ che implica, per l'iniettività di $\mathbf{f}$, $\mathbf{x} = \mathbf{z}$. Dunque ogni valore limite di $\{\mathbf{x}_k\}$ coincide con $\mathbf{x}$, ovvero $\mathbf{x}_k \to \mathbf{x}$. □

La definizione 1.2 di funzione uniformemente continua si estende pari pari al caso di più variabili; diremo infatti che $\mathbf{f} : \mathbb{R}^n \supseteq X \to \mathbb{R}^m$ *è uniformemente*

continua in X se, $\forall\varepsilon > 0$, $\exists\delta > 0$ ($\delta = \delta(\varepsilon)$) tale che, per ogni coppia di punti $\mathbf{x}, \mathbf{z} \in X$ con $\|\mathbf{x} - \mathbf{z}\| < \delta$ risulti $\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{z})\| < \varepsilon$.

▪ **Teorema 2.7** - (Cantor–Heine) - *Sia $K \subset \mathbb{R}^n$ compatto; sia $\mathbf{f} : K \to \mathbb{R}^m$ continua. Allora $\mathbf{f}$ è uniformemente continua.*

Dimostrazione - Per assurdo: $\mathbf{f}$ non sia uniformemente continua; allora esiste $\varepsilon_0 > 0$ tale che, per ogni $\delta > 0$, ci sono punti $\mathbf{x}_\delta$ e $\mathbf{z}_\delta$ (dipendenti da δ) tali che:

$$\|\mathbf{x}_\delta - \mathbf{z}_\delta\| < \delta \quad \text{e} \quad \|\mathbf{f}(\mathbf{x}_\delta) - \mathbf{f}(\mathbf{z}_\delta)\| \geq \varepsilon_0 . \tag{2.3}$$

Diamo a δ i valori $1, 1/2, 1/3, \cdots, 1/k$, e siano $\mathbf{x}_k$ e $\mathbf{z}_k$ i corrispondenti punti con la proprietà (2.3). Per la compattezza di K, dalla successione $\{\mathbf{x}_k\}$ si può estrarre una sottosuccessione, sia $\{\mathbf{x}_{k_l}\}$, convergente a un certo $\mathbf{x} \in K$. Ma essendo

$$\|\mathbf{z}_{k_l} - \mathbf{x}\| \leq \|\mathbf{z}_{k_l} - \mathbf{x}_{k_l}\| + \|\mathbf{x}_{k_l} - \mathbf{x}\| \leq \frac{1}{k_l} + \|\mathbf{x}_{k_l} - \mathbf{x}\| \to 0$$

per $k_l \to \infty$, risulta che anche $\{\mathbf{z}_{k_l}\}$ converge a $\mathbf{x}$.

Ma ancora si ha:

$$\|\mathbf{f}(\mathbf{x}_{k_l}) - \mathbf{f}(\mathbf{z}_{k_l})\| \leq \|\mathbf{f}(\mathbf{x}_{k_l}) - \mathbf{f}(\mathbf{x})\| + \|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{z}_{k_l})\|$$

e, per la continuità di $\mathbf{f}$, il secondo membro tende a 0 quando $k_l \to \infty$; pertanto il primo membro tende a 0, contraddicendo l'ipotesi (2.3). □

2.3 Funzioni continue su un connesso

▪ **Teorema 2.8** - *Sia $E \subseteq \mathbb{R}^n$ connesso; sia $\mathbf{f} : E \to \mathbb{R}^m$ continua. Allora $\mathbf{f}(E)$ è connesso.*

Conviene far precedere alla dimostrazione del teorema il seguente Lemma, che caratterizza gli insiemi connessi per mezzo della topologia indotta.

Lemma 2.9 - *$E \subseteq \mathbb{R}^n$ è connesso se e solo se gli unici insiemi contemporaneamente aperti e chiusi in E sono $\emptyset$ ed E stesso.*

Dimostrazione - Sia E non connesso; allora $E = A \cup B$ dove A e B sono non vuoti e separati. Allora A è aperto in E poiché $A = (\mathbb{R}^n \setminus \overline{B}) \cap E$ (intersezione di un aperto di $\mathbb{R}^n$ con E); analogamente B è aperto in E.

Si ha anche $B = \overline{B} \cap E$; infatti se $\mathbf{x}_0 \in \overline{B} \cap E$ allora $\mathbf{x}_0 \in E$ e $\mathbf{x}_0 \notin A$ essendo $A \cap \overline{B} = \emptyset$; deve dunque essere $\mathbf{x}_0 \in B$; ciò mostra che B è chiuso in E. Analogamente si mostra che A è chiuso in E. Ma allora A e B sono contemporaneamente chiusi e aperti in E, diversi da $\emptyset$ e da E.

Viceversa, se $A (\neq \emptyset$ e $\neq E)$ è aperto e chiuso in E, posto $B = E \setminus A$ si ha: $A \cup B = E$, $A \cap B = \emptyset$ ed è anche facile verificare che $A \cap \overline{B} = \emptyset$ e $\overline{A} \cap B = \emptyset$. □

Dimostrazione del teorema 2.8 - Ancora una dimostrazione per assurdo. Sia $\mathbf{f}(E)$ non connesso; allora esiste $A \subset \mathbf{f}(E)$, A aperto e chiuso in $\mathbf{f}(E)$, $A \neq \emptyset$ e $A \neq f(E)$.

Dunque $A = U \cap \mathbf{f}(E)$ e $A = V \cap f(E)$ dove U e V sono un aperto e un chiuso di $\mathbb{R}^m$ rispettivamente. Ma allora $\mathbf{f}^{-1}(A)$ è aperto e chiuso in E (Teorema 2.1 e Corollario 2.2), $\mathbf{f}^{-1}(A) \neq \emptyset$ e $\mathbf{f}^{-1}(A) \neq E$. Ne segue, per il Lemma 2.9, che E non è connesso, contro l'ipotesi. □

In particolare notiamo che, se $f : \mathbb{R} \to \mathbb{R}$ è continua, allora l'immagine di un intervallo è ancora un intervallo (limitato o no). Dal teorema ora dimostrato si ricava subito il teorema 1.4 (teorema degli zeri di una funzione continua): infatti l'immagine $f([a,b])$ di $[a,b]$ è ancora un intervallo contenente un punto < 0 ($f(a)$ per esempio) e un altro ($f(b)$) > 0: perciò deve contenere anche lo 0, il che significa che esiste $\xi \in [a,b]$ che ha come immagine lo 0. Anche il Corollario 1.5 può dedursi direttamente dal teorema 2.8.

Dal teorema 2.8 deduciamo infine il seguente teorema di continuità per la funzione inversa di una funzione da $\mathbb{R}$ in $\mathbb{R}$

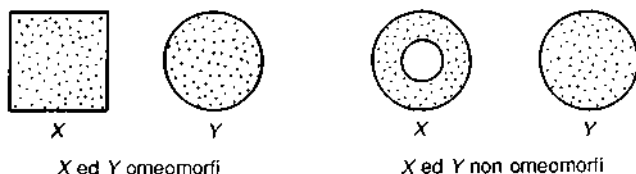
■ **Teorema 2.10** - *Sia $I \subseteq \mathbb{R}$ un intervallo e $f : I \to \mathbb{R}$ continua e iniettiva. Allora $f^{-1} : f(I) \to I$ è continua.*

Dimostrazione - Sia y un punto interno di $f(I)$ e $[c,d]$ un intervallo contenuto in $f(I)$ e contenente y al suo interno.

Siano $a, b \in I$ tali che $f(a) = c$ e $f(b) = d$; supponiamo $a < b$. Allora l'immagine di $[a,b]$, compatto e connesso, è pure compatta e connessa e siccome $f(a) = c < y$ e $f(b) = d > y$, il punto y appartiene a $f([a,b])$; ma f^{-1} è continua su $f([a,b])$ per il teorema 2.6 e perciò f^{-1} è continua in y. Per l'arbitrarietà di y, f^{-1} risulta continua in tutti i punti interni di $f(I)$. Chiaramente la dimostrazione vale anche se y è un estremo di $f(I)$. Perciò f^{-1} è continua in tutti i punti di $f(I)$. □

Concludiamo questa sezione già ricca di concetti topologici con la seguente definizione. Se due insiemi X e Y possono essere posti in corrispondenza biunivoca tramite una funzione f *bicontinua* (tale cioè che sia f sia f^{-1} siano continue) essi si dicono **omeomorfi** e f (o f^{-1}) si chiama **omeomorfismo** tra X e Y.

Gli omeomorfismi costituiscono la realizzazione matematica del concetto intuitivo di *deformazione continua*: ad esempio è intuitivo che se X è un quadrato chiuso (aperto) e Y è un cerchio chiuso (aperto), entrambi in $\mathbb{R}^2$, allora X e Y sono omeomorfi; non sono omeomorfi invece un cerchio ed una corona circolare.

Fig. 5.5 X ed Y omeomorfi — X ed Y non omeomorfi

La *topologia* è la branca della matematica che studia quei concetti che si conservano in un omeomorfismo. Aperti, chiusi, intorni, punti di accumulazione sono

esempi di concetti topologici in quanto vengono trasformati in aperti, chiusi, intorni, punti di accumulazione (rispettivamente), eventualmente nella topologia indotta.

Non sono invece concetti topologici (in $\mathbb{R}^n$) quelli di distanza, parallelismo, perpendicolarità che vengono "distrutti" da una deformazione continua.

3. FUNZIONI ELEMENTARI

3.1 Funzioni razionali intere. Polinomi

Si dice *funzione razionale intera* (di variabile reale) una funzione $P : \mathbb{R} \to \mathbb{R}$ del tipo

$$P(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \tag{3.1}$$

dove $a_0, a_1, \cdots, a_n \in \mathbb{R}$ sono detti *coefficienti* di P. Se, nella (3.1), $a_n \neq 0$, l'intero naturale n si chiama *grado* di P (si scrive $n = \text{gr}\, P$); se $n = 0$ la funzione è una costante.

Due funzioni intere si dicono *uguali* se hanno lo stesso grado ed hanno ordinatamente uguali i coefficienti.

Poiché le potenze sono funzioni continue su tutto $\mathbb{R}$, così sono continue su tutto $\mathbb{R}$ le funzioni razionali intere. Inoltre sia la somma di due funzioni di questo tipo che il prodotto di una tale funzione per un numero sono ancora funzioni dello stesso tipo: il che significa che l'insieme delle funzioni razionali intere forma uno spazio lineare, sottospazio di $C^0(\mathbb{R})$ (in esso lo $\mathbf{0}$ è la funzione identicamente nulla, e l'opposto di $P(x)$ è $-P(x)$).

Inoltre si può anche eseguire il prodotto, con le usuali regole dell'algebra, di due funzioni intere $P(x)$ e $Q(x)$, e il risultato è ancora una funzione intera di grado uguale alla somma dei gradi dei fattori; siano:

$$P(x) = a_0 + a_1 x + \cdots + a_n x^n$$

$$Q(x) = b_0 + b_1 x + \cdots + b_m x^m \ .$$

Si trova

$$P(x)Q(x) = c_0 + c_1 x + \ldots + c_{n+m} x^{n+m}$$

dove

$$\begin{aligned} c_0 &= a_0 b_0 \\ c_1 &= a_0 b_1 + a_1 b_0 \\ &\vdots \\ c_k &= a_0 b_k + a_1 b_{k-1} + \cdots + a_{k-1} b_1 + a_k b_0 \\ &\vdots \end{aligned}$$

Si osserverà che l'insieme di queste funzioni, rispetto alle operazioni di somma e prodotto, ha proprietà algebriche somiglianti a quelle dell'insieme $\mathbb{Z}$ degli interi relativi; più precisamente, rispetto alla somma esso forma un gruppo commutativo, e inoltre è definito un prodotto per il quale valgono la proprietà associativa e commutativa e rispetto al quale esiste l'unità (la funzione costante 1); inoltre vale la proprietà distributiva: un tale insieme si dice **anello** (commutativo, con unità). Questa considerazione conduce ad estendere notevolmente il concetto di funzione intera, considerandolo puramente sotto l'aspetto algebrico.

Definizione 3.1 - *Sia* Λ *un campo* (per esempio $\mathbb{Q}, \mathbb{R}, \mathbb{C}$). *Si dice* **polinomio** *in* x *a coefficienti in* Λ *una espressione del tipo*

$$P = a_0 + a_1 x + \cdots + a_n x^n \quad , \; n \in \mathbb{N} \tag{3.2}$$

dove $a_0, a_1, \cdots, a_n$ *sono elementi di* Λ.

La x, nella (3.2), si dice *indeterminata* e svolge un ruolo puramente simbolico, cioè non è più necessario che essa sia una variabile numerica, ma semplicemente un oggetto per cui sono definite le potenze (per esempio, una matrice quadrata). Ai polinomi si estendono in maniera ovvia le definizioni e le operazioni sopra riportate per le funzioni razionali intere. Parleremo perciò di anello dei polinomi a coefficienti razionali, oppure reali, oppure complessi. Osserviamo anche che spesso, ma impropriamente, si chiama polinomio la funzione che viene ad esso associata quando x è una variabile numerica.

La somiglianza dei polinomi con i numeri relativi si estende alla possibilità di eseguire la divisione secondo le potenze decrescenti. Vale infatti il seguente

■ **Teorema 3.1** - *Siano* A *e* B *due polinomi, con* $B \neq 0$. *Esistono due polinomi* Q *e* R, *univocamente determinati, tali che*

$$A = BQ + R \quad , \quad \mathrm{gr} R < \mathrm{gr} B \, . \tag{3.3}$$

Il polinomio Q è detto *quoziente*, R *resto* della divisione di A per B.

La dimostrazione del teorema 3.1 è accennata negli esercizi 3.1 (unicità) e 3.2.

Si osservi che se $\mathrm{gr}A < \mathrm{gr}B$, allora $Q = 0$ e $R = A$. Se $\mathrm{gr}A = \mathrm{gr}B$ dovrà essere $\mathrm{gr}Q = 0$, cioè $Q =$ costante. Se $\mathrm{gr}B = 1$, dovrà essere $\mathrm{gr}R = 0$ e cioè $R =$costante.

Se nella divisione risulta $R = 0$ si dice che B *divide* A (B è un *divisore* di A). Si dice che A (di grado ≥ 1) è *irriducibile* se non ammette divisori B con $0 < \mathrm{gr}B < \mathrm{gr}A$. Il fatto di ammettere divisori oppure no dipende naturalmente dal campo sul quale il polinomio è definito.

Esempi

3.1. $1 + x^2$ su $\mathbb{R}$ è irriducibile, mentre su $\mathbb{C}$ ha dei divisori (per esempio, $1 + ix$).

3.2. Un polinomio di 2° grado: $ax^2 + bx + c$ con discriminante < 0 (cioè $b^2 - 4ac < 0$) è irriducibile su $\mathbb{R}$.

3.3. $2 - x^2$ è irriducibile su $\mathbb{Q}$, mentre su $\mathbb{R}$ ha dei divisori (per esempio, $\sqrt{2} - x$).

3.4. I polinomi di 1° grado sono tutti irriducibili su qualsiasi campo.

Definizione 3.2 - *Se il polinomio $x - c$ ($c \in \Lambda$) è un divisore di A, si dice che c è una* **radice** *(o uno* **zero***) di A; se $(x - c)^m$ (m intero ≥ 1) è un divisore di A, ma $(x - c)^{m+1}$ non lo è, diremo che c è una radice di molteplicità (o ordine) m.*

Se c è una radice di A di molteplicità m, si potrà scrivere allora

$$A(x) = (x - c)^m Q(x)$$

dove Q non è il polinomio zero e non ammette c come radice.

Se $c_1, c_2, \cdots, c_k$ sono radici di A di molteplicità rispettiva $m_1, m_2, \cdots, m_k$ si avrà

$$A(x) = (x - c_1)^{m_1}(x - c_2)^{m_2} \cdots (x - c_k)^{m_k} Q(x)$$

$$\text{gr}A = m_1 + m_2 + \cdots + m_k + \text{gr}Q$$

e Q non è il polinomio zero e non ammette alcuno dei numeri $c_1, \cdots, c_k$ come radice.

Perciò dovrà essere: $m_1 + m_2 + \cdots + m_k \leq \text{gr}A$. Da ciò segue immediatamente la seguente importante

Proposizione 3.2 - *Un polinomio di grado n non può avere più di n radici.*

Pertanto, se un polinomio $A(x)$ si annulla per più di n valori distinti della x, esso è identicamente nullo. Vale perciò il seguente

Principio di identità dei polinomi: *se due polinomi $P(x)$ e $Q(x)$ di grado $\leq n$ assumono valori uguali per almeno $n + 1$ valori distinti della x, essi sono uguali, cioè $P(x) = Q(x)$ per ogni x.*

Consideriamo ora i polinomi sul campo complesso $\mathbb{C}$. Abbiamo già ricordato in 2.4.3, nella forma di un suo corollario, il teorema fondamentale dell'Algebra (di Gauss) e cioè:

▪ **Teorema 3.3** - *Ogni polinomio P sul campo complesso di grado ≥ 1 ammette almeno una radice.*

Il teorema ricordato in 2.4.3 si deduce immediatamente dal precedente osservando che, se $\text{gr}P = n$ e c_1 è uno zero di P, potremo scrivere

$$P(x) = (x - c_1)Q_1(x)$$

con $\text{gr}Q_1 = n - 1$; se c_2 (eventualmente coincidente con c_1) è uno zero di Q_1, potremo scrivere

$$P(x) = (x - c_1)(x - c_2)Q_2(x)$$

con gr$Q_2 = n-2$, e così via finché si perviene ad un polnomio Q_n di grado 0, cioè ad una costante. Allora, se $P(x) = a_0 + a_1x + \cdots + a_nx^n$ è un polinomio di grado n sul campo complesso, e $c_1, c_2, \cdots, c_k$ sono le sue radici *distinte* di molteplicità rispettiva $m_1, m_2, \cdots, m_k$, risulta precisamente:

$$m_1 + m_2 + \cdots + m_k = n \tag{3.4}$$

e, per ogni $x \in \mathbb{C}$,

$$P(x) = a_n(x-c_1)^{m_1}(x-c_2)^{m_2}\cdots(x-c_k)^{m_k} \ . \tag{3.5}$$

Si osservi che, in $\mathbb{C}$, i soli polinomi irriducibili sono quelli di primo grado. La (3.5) ci fornisce la decomposizione di un polinomio P sul campo $\mathbb{C}$ in fattori irriducibili.

Sia ora $P = a_0 + a_1x + \cdots + a_nx^n$; si dice polinomio coniugato di P, e si indica con $\overline{P}$, il polinomio $\overline{a}_0 + \overline{a}_1x + \cdots + \overline{a}_nx^n$ i cui coefficienti sono i coniugati dei coefficienti di P. È evidente che, se $c \in \mathbb{C}$ è uno zero di P (di molteplicità m), allora $\overline{c}$ è uno zero di $\overline{P}$ (con la stessa molteplicità) e viceversa.

Supponiamo ora che i coefficienti di P siano reali e perciò $\overline{P} = P$. Allora, se $c \in \mathbb{C}$ è una radice di P, anche $\overline{c}$ lo è.

Cerchiamo una decomposizione come la (3.5) per un polinomio P (di grado n) sul campo reale. Siano $c_1, c_2, \cdots, c_k$ le radici reali e distinte di P di molteplicità $m_1, m_2, \cdots, m_k$ e siano $\gamma_1, \overline{\gamma}_1, \gamma_2, \overline{\gamma}_2, \cdots, \gamma_h, \overline{\gamma}_h$ le radici complesse distinte di P di molteplicità $\mu_1, \mu_2, \cdots, \mu_h$; allora sarà

$$m_1 + \cdots + m_k + 2\mu_1 + \cdots + 2\mu_h = n \tag{3.6}$$

e, per ogni $x \in \mathbb{R}$

$$\begin{aligned} P(x) = a_n(x-c_1)^{m_1}\cdots(x-c_k)^{m_k}(x-\gamma_1)^{\mu_1}(x-\overline{\gamma}_1)^{\mu_1}\cdots \\ \cdots(x-\gamma_h)^{\mu_h}(x-\overline{\gamma}_h)^{\mu_h} \ . \end{aligned} \tag{3.7}$$

Ma osserviamo che

$$(x-\gamma_i)(x-\overline{\gamma}_i) = x^2 + p_ix + q_i \qquad (i = 1, 2, \cdots, h)$$

con

$$p_i = -2\text{Re}\,\gamma_i \ , \ q_i = |\gamma_i|^2 \ ,$$

e poiché $p_i^2 - 4q_i = -4(\text{Im}\,\gamma_i)^2 < 0$ i polinomi $x^2 + p_ix + q_i$ sono irriducibili sul campo reale. La (3.7) si scrive allora

$$\begin{aligned} P(x) = &a_n(x-c_1)^{m_1}\cdots(x-c_k)^{m_k}(x^2+p_1x+q_1)^{\mu_1}\cdots \\ &\cdots(x^2+p_hx+q_h)^{\mu_h} \end{aligned} \tag{3.8}$$

che fornisce la cercata decomposizione di P in fattori irriducibili. Sul campo reale gli unici polinomi irriducibili sono dunque quelli di 1° grado e quelli di 2° grado con discriminante negativo.

3.2 Funzioni razionali fratte

Come dall'insieme degli interi relativi $\mathbb{Z}$ si passa all'insieme dei numeri razionali $\mathbb{Q}$, così, con procedimento analogo, dall'insieme dei polinomi si può costruire un altro insieme, quello delle frazioni razionali.

In breve, si consideri l'insieme delle coppie (P, Q) di polinomi (su un campo Λ) con Q diverso dal polinomio zero; si introduca una equivalenza ponendo

$$(P_1, Q_1) \approx (P_2, Q_2) \quad \text{se risulta} \quad P_1 Q_2 = P_2 Q_1 \ .$$

Le classi di equivalenza si scriveranno semplicemente come frazioni: P/Q sta per $[(P, Q)]$: classe di equivalenza di cui un rappresentante è la coppia (P, Q). Nella scelta dei rappresentanti P e Q conveniamo di prendere due polinomi *primi tra loro* (*cioè privi di divisori comuni*): questi polinomi sono così determinati in modo unico, a meno di una costante moltiplicativa. Per esempio, essendo

$$(x^2 - 1, x^2 - 3x + 2) \approx (x + 1, x - 2)$$

la frazione rappresentata da queste coppie sarà indicata con

$$\frac{x+1}{x-2} \quad \text{e non} \quad \frac{x^2-1}{x^2-3x+2} \ .$$

Su queste frazioni si definiscono le operazioni di somma e prodotto con le solite regole del calcolo algebrico:

$$\frac{P}{Q} + \frac{R}{S} := \frac{PS + RQ}{QS}$$

$$\frac{P}{Q} \cdot \frac{R}{S} := \frac{PR}{QS}$$

(bisogna verificare che, cambiando i rappresentanti delle classi a primo membro, non cambia la classe a secondo membro).

L'insieme così costruito ha le proprietà di corpo commutativo: si chiama corpo delle *frazioni razionali*.

Se consideriamo i polinomi sul campo reale, ad ogni frazione razionale risulta associata una funzione reale di variabile reale così definita

$$x \mapsto R(x) = \frac{P(x)}{Q(x)}$$

(con $P(x)$ e $Q(x)$ funzioni razionali intere). $R(x)$ è definita su $\mathbb{R}$ privato degli zeri reali di Q (cioè è definita sull'unione di un numero finito di intervalli aperti) ed è

continua, poiché rapporto di due funzioni continue. Gli zeri di Q si dicono *poli* di R (di ordine m, se m è la molteplicità dello zero di Q). Analoghe considerazioni valgono per le funzioni razionali sul campo complesso. I grafici di due funzioni razionali molto semplici sono illustrati nei diagrammi 10 e 11.

Data una frazione $R = P/Q$ (scritta in forma ridotta, cioè con P e Q primi fra loro), se $\mathrm{gr}P \geq \mathrm{gr}Q$ possiamo anzitutto procedere alla divisione di P per Q ottenendo

$$P = SQ + T \quad \text{con} \quad \mathrm{gr}T < \mathrm{gr}Q$$

e quindi possiamo scrivere

$$R = S + T/Q \; .$$

Supponiamo ora i polinomi definiti sul campo complesso; allora, per la (3.5), possiamo scrivere

$$Q(x) = q(x - c_1)^{m_1}(x - c_2)^{m_2} \cdots (x - c_k)^{m_k} \tag{3.9}$$

essendo $c_1, c_2, \cdots, c_k$ le radici distinte di Q e $m_1, m_2, \cdots, m_k$ le loro rispettive molteplicità. Il seguente teorema, che illustreremo solo con esempi, ci permette di rappresentare la frazione P/Q mediante elementi più semplici.

▪ **Teorema 3.4** - *Siano* T, Q *due polinomi su* $\mathbb{C}$, *con* $\mathrm{gr}T < \mathrm{gr}Q$ *e valga la* (3.9). *Allora la frazione* T/Q *si può rappresentare in modo unico nella forma:*

$$\begin{aligned}\frac{T(x)}{Q(x)} =& \frac{c_{11}}{(x - c_1)} + \frac{c_{12}}{(x - c_1)^2} + \cdots + \frac{c_{1m_1}}{(x - c_1)^{m_1}} + \\ & \frac{c_{21}}{x - c_2} + \frac{c_{22}}{(x - c_2)^2} + \cdots + \frac{c_{2m_2}}{(x - c_2)^{m_2}} + \\ & \cdots + \frac{c_{k1}}{x - c_k} + \frac{c_{k2}}{(x - c_k)^2} + \cdots + \frac{c_{km_k}}{(x - c_k)^{m_k}} \; .\end{aligned} \tag{3.10}$$

La (3.10) si dice rappresentazione in *fratti semplici* della frazione T/Q.

Esempio **3.5** - Nella frazione

$$R(x) = \frac{P(x)}{Q(x)} = \frac{x^2 + 2x + 2}{1 + x^2}$$

risulta $\mathrm{gr}P = \mathrm{gr}Q$. Allora, dividendo P per Q, si otterrà $P = SQ + T$ con S costante e $\mathrm{gr}T < \mathrm{gr}Q$. Poi, essendo $\pm i$ le radici, semplici e distinte, di Q, in base al teorema 3.4 dovrà valere la rappresentazione

$$\frac{T(x)}{Q(x)} = \frac{c_1}{x - i} + \frac{c_2}{x + i} \; .$$

In conclusione dovrà risultare

$$R(x) = c_0 + \frac{c_1}{x-i} + \frac{c_2}{x+i} \,.$$

Le costanti c_0, c_1, c_2 si determinano facilmente riducendo ambo i membri della precedente equazione allo stesso denominatore e poi uguagliando i coefficienti delle potenze simili dei numeratori.

Si trova

$$c_0 = 1 \qquad c_1 = 1 - \frac{i}{2} \qquad c_2 = 1 + \frac{i}{2}$$

Analogamente, sul campo reale, varrà una decomposizione fondata sulla possibilità di rappresentare un polinomio nella forma (3.8).

■ **Teorema 3.5** - *Siano* T, Q *due polinomi su* $\mathbb{R}$, *con* $\operatorname{gr}T < \operatorname{gr}Q$ *e sia*

$$Q(x) = q(x-c_1)^{m_1} \cdots (x-c_k)^{m_k} (x^2 + p_1 x + q_1)^{\mu_1} \cdots (x^2 + p_h x + q_h)^{\mu_h}$$

la rappresentazione di Q *mediante elementi irriducibili. Allora esiste una rappresentazione unica di* T/Q *della forma:*

$$\frac{T(x)}{Q(x)} = \frac{c_{11}}{x-c_1} + \cdots + \frac{c_{1m_1}}{(x-c_1)^{m_1}} + \cdots + \frac{c_{k1}}{x-c_k} + \cdots + \frac{c_{km_k}}{(x-c_k)^{m_k}} +$$

$$+ \frac{p_{11}x + q_{11}}{x^2 + p_1 x + q_1} + \cdots + \frac{p_{1\mu_1}x + q_{1\mu_1}}{(x^2 + p_1 x + q_1)^{\mu_1}} + \cdots +$$

$$+ \frac{p_{h_1}x + q_{h_1}}{x^2 + p_h x + q_h} + \cdots + \frac{p_{h\mu_h}x + q_{h\mu_h}}{(x^2 + p_h x + q_h)^{\mu_h}} \,.$$

Esempi

3.6. $$\frac{1}{x^2(x^2+x+1)} = \frac{a}{x} + \frac{b}{x^2} + \frac{cx+d}{x^2+x+1}.$$

Riducendo allo stesso denominatore e uguagliando i coefficienti delle potenze simili si trova:

$$a = -1 \qquad b = 1 \qquad c = 1 \qquad d = 0 \,.$$

3.7. $$\frac{1}{(x+1)(1+x^2)^2} = \frac{a}{x+1} + \frac{bx+c}{x^2+1} + \frac{dx+e}{(x^2+1)^2}$$

Si trova

$$a = \frac{1}{4} \qquad b = -\frac{1}{4} \qquad c = \frac{1}{4} \qquad d = -\frac{1}{2} \qquad e = \frac{1}{2} \,.$$

Esercizi

1. Dimostrare l'unicità affermata nel teorema 3.1 (assumere che esistano due rappresentazioni: $A = BQ_1 + R_1$, $A = BQ_2 + R_2$, con $\mathrm{gr}R_1 < \mathrm{gr}B$ e $\mathrm{gr}R_2 < \mathrm{gr}B$...).

2. *Regola per la divisione dei polinomi*

Si voglia eseguire la divisione tra

$$A(x) = 3x^3 - 2x + 1 \ .$$

e

$$B(x) = x^2 - 4x \ .$$

Si divide $3x^3$ per x^2 ottenendo $3x$ (1° monomio del quoziente). Si moltiplica $B(x)$ per $3x$ ottenendo $3x^3 - 12x^2$ che si sottrae da $B(x)$ ricavando $12x^2 - 2x + 1$. Conviene eseguire i calcoli secondo lo schema illustrato qui di seguito:

$$\begin{array}{lll|l}
3x^3 & & -2x+1 & x^2-4x \\ \cline{4-4}
3x^3-12x^2 & & & 3x \\ \cline{1-3}
 & +12x^2 & -2x+1 &
\end{array}$$

Si ripete l'algoritmo con i polinomi $12x^2 - 2x + 1$ e $x^2 - 4x$.

Dividendo $12x^2$ per x^2 si ottiene il secondo monomio del quoziente: 12. Si moltiplica ora 12 per $x^2 - 4$ ed il risultato si sottrae a $12x^2 - 2x + 1$; lo schema completo è il seguente:

$$\begin{array}{lll|l}
3x^3 & & -2x+1 & x^2-4x \\ \cline{4-4}
3x^3-12x^2 & & & 3x+12 \\ \cline{1-3}
 & +12x^2 & -\ 2x+1 & \\
 & +12x^2 & -48x & \\ \cline{1-3}
 & & +46x+1 &
\end{array}$$

Essendo $46x + 1$ di grado inferiore a $\mathrm{gr}B$, la divisione è finita; si ha: $Q(x) = 3x + 12$, $R(x) = 46x + 1$.

3. Decomporre in $\mathbb{C}$ le seguenti frazioni in fratti semplici:

a) $\dfrac{1}{1+x^4}$

b) $\dfrac{1}{(1+x^2)^2}$

c) $\dfrac{x^2+2}{(x^2+3)^3}$

d) $\dfrac{2x^3+3}{x(x+1)(x^2+2)}$

4. Decomporre in $\mathbb{R}$ le seguenti frazioni in fratti semplici:

a) $\dfrac{1}{1+x^4}$

b) $\frac{2x^2+5x-1}{x^3+x^2-2x}$

c) $\frac{x+1}{x^3-1}$

d) $\frac{1}{x^3(1+x)^2}$.

3.3 Funzioni algebriche

Le funzioni razionali intere e fratte si ottengono operando con la variabile x e le costanti un numero finito di volte mediante le quattro operazioni algebriche elementari. Se ora facciamo intervenire anche l'operazione di estrazione di radice n–esima, definita in 2.3.1 e consideriamo la composizione, con un numero finito di fattori, di funzioni razionali con le funzioni $x^{1/n}$ $(n = 1, 2, \cdots)$ otteniamo le *funzioni algebriche*. Anch'esse, essendo composte con funzioni continue, sono continue in tutto l'insieme di definizione.

I grafici delle funzioni algebriche più semplici sono illustrati nei diagrammi n. 7, 8, 12, 13.

È interessante osservare che, in virtù della definizione data, una funzione algebrica $y = f(x)$ verifica una equazione del tipo

$$P_0(x) + P_1(x)y + \cdots + P_n(x)y^n = 0 \tag{3.11}$$

per un certo $n \in \mathbb{N}$, dove $P_0, P_1, \cdots, P_n$ sono polinomi (una tale equazione viene detta appunto *equazione algebrica*).

Esempi

3.8. $y = \frac{1}{\sqrt{x}+1}$ verifica l'equazione del 2° ordine: $(1-x)y^2 - 2y + 1 = 0$.

3.9. $y = 3 + \sqrt[3]{1-x^2}$ verifica l'equazione del 3° ordine: $(27-x)y^3 - 27xy^2 + 9x^2y - x^3 = 0$.

3.10. $y = \sqrt{x + \sqrt{x + \sqrt{x}}}$ verifica l'equazione del 4° ordine: $y^4 - 4xy^2 - 4xy - x = 0$.

Si può dimostrare che una equazione algebrica di grado non superiore al quarto è sempre risolubile per mezzo di radicali (le formule esplicite risolutive per le equazioni di terzo grado sono state ricavate a pag. 115), e cioè ogni soluzione della (3.11) con $n \leq 4$ è una funzione algebrica.

È stato anche dimostrato (ricordiamo, in particolar modo, Ruffini, Abel, Galois) che, in generale, le equazioni algebriche di grado superiore al quarto non sono risolubili per radicali. Ciò ha condotto ad estendere il concetto di funzione alge-

brica, includendovi anche quelle funzioni che sono definite "implicitamente" dalle equazioni algebriche di grado superiore al quarto (il concetto di funzione definita implicitamente da un'equazione sarà sviluppato più avanti); ma non intendiamo entrare più a fondo in tali questioni.

3.4 Esponenziali e logaritmi

La funzione $\exp_a : x \mapsto a^x$, con a reale > 0, $a \neq 1$, è una delle più importanti dell'Analisi, ed è stata già considerata più volte. Ricordiamone qui le proprietà principali:

$$\mathrm{dom}(\exp_a) = \mathbb{R} \quad , \quad \mathrm{im}(\exp_a) = \mathbb{R}_+$$

$$a^x \quad \text{è continua} \quad \forall\, x \in \mathbb{R}$$

$$a^x \quad \text{è monotona} \begin{cases} \text{crescente} & \text{per } a > 1 \\ \text{decrescente} & \text{per } a < 1 \end{cases}$$

$$\lim_{x \to -\infty} a^x = \begin{cases} 0 & \text{per } a > 1 \\ +\infty & \text{per } a < 1 \end{cases}$$

$$\lim_{x \to +\infty} a^x = \begin{cases} +\infty & \text{per } a > 1 \\ 0 & \text{per } a < 1 \,. \end{cases}$$

I grafici di queste funzioni, per diversi valori di a, sono riportati nei diagrammi n. 14 e 15.

Inoltre sappiamo che:

$$a^{x+y} = a^x a^y \quad \text{per ogni} \quad x, y \in \mathbb{R} \tag{3.12}$$

da cui segue, in particolare, che $a^0 = 1$. La (3.12) mostra che la funzione $\exp_a : \mathbb{R} \to \mathbb{R}_+$ trasforma l'operazione $+$ in $\mathbb{R}$ nell'operazione $\cdot$ in $\mathbb{R}_+$, manda l'elemento neutro della somma (lo zero) in $\mathbb{R}$ nell'elemento neutro del prodotto (l'unità) in $\mathbb{R}_+$; in altre parole, l'insieme $\mathbb{R}$, considerato come gruppo commutativo rispetto alla somma (lo indicheremo $(\mathbb{R}, +)$) viene trasformato dall'applicazione $\exp_a$ nell'insieme $\mathbb{R}_+$, considerato come gruppo commutativo rispetto al prodotto (sarà indicato $(\mathbb{R}_+, \cdot)$). È importante notare che, tra le funzioni continue, l'esponenziale è l'unica a realizzare una simile trasformazione, cioè l'equazione (3.12) caratterizza le funzioni esponenziali, nel senso indicato nel seguente teorema.

▪ **Teorema 3.6** - *Sia $f : \mathbb{R} \to \mathbb{R}$ una funzione continua (non costante) per cui valga l'equazione*

$$f(x + y) = f(x) f(y) \qquad \forall x, y \in \mathbb{R} \,. \tag{3.13}$$

Allora $f(x) = a^x$ per qualche $a > 0$.

Dimostrazione - Consiste in alcune semplici deduzioni dalla (3.13).

i) Ponendo $x = y = 0$ nella (3.13) si ha: $f(0) = f(0)f(0)$ da cui $f(0) = 1$ oppure $f(0) = 0$.

ii) Se $f(0) = 0$, ponendo $y = 0$ nella (3.13) si ha: $f(x) = f(x)f(0) = 0$ e perciò $f(x) = 0$ per ogni $x \in \mathbb{R}$.

iii) Sia allora $f(0) = 1$; ponendo $y = -x$ nella (3.13) si ha: $1 = f(x-x) = f(x)f(-x)$ e cioè $f(x) \neq 0$, $\forall x \in \mathbb{R}$; essendo continua e sempre diversa da zero deve essere o sempre positiva o sempre negativa; essendo $f(0) = 1$ sarà sempre > 0.

iv) Poniamo $f(1) = a > 0$. Dalla (3.13) otteniamo:

$$f(2) = f(1+1) = f(1)f(1) = a^2 \;\; ; \;\; f(3) = f(2)f(1) = a^3$$

e, per induzione, $f(n) = a^n$, $\forall n \in \mathbb{N}$.

Essendo $1 = f(n-n) = f(n)f(-n)$ ricaviamo $f(-n) = a^{-n}$. Sappiamo così calcolare f su $\mathbb{Z}$.

v) Calcoliamo ora f sui razionali. Sia $r = 1/n$. Si ha, usando ripetutamente la (3.13)

$$a = f(1) = f(\underbrace{\frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n}}_{n \text{ volte}}) = \left[f\left(\frac{1}{n}\right)\right]^n$$

Dunque $f(1/n) = a^{1/n}$

Inoltre

$$f\left(\frac{m}{n}\right) = f(\underbrace{\frac{1}{n} + \cdots + \frac{1}{n}}_{m \text{ volte}}) = \left[f\left(\frac{1}{n}\right)\right]^m = a^{m/n} .$$

vi) Finalmente, se $x \in \mathbb{R}$ ed $\{r_n\}$ è una successione di razionali che tende ad x, usando la continuità di f si ha:

$$f(x) = f\left(\lim_{n\to+\infty} r_n\right) = \lim_{n\to+\infty} f(r_n) = \lim_{n\to+\infty} a^{r_n} = a^x .$$

Il teorema è così dimostrato. □

Una equazione come la (3.13), in cui compare come incognita la funzione f valutata in due o più punti, viene detta *equazione funzionale*. Incontreremo tra poco un altro esempio di equazione di questo tipo.

La funzione esponenziale $\exp_a : \mathbb{R} \to \mathbb{R}_+$, essendo biiettiva (perché monotona) è invertibile e la sua inversa $\exp_a^{-1} : \mathbb{R}_+ \to \mathbb{R}$ prende il nome di *logaritmo* e si indica con $\log_a$.

Le proprietà principali della funzione logaritmo, già note allo studente, si possono facilmente dedurre da quelle dell'esponenziale; in particolare essa è continua,

essendo l'inversa di una funzione continua definita su un intervallo. Inoltre vale l'uguaglianza:

$$\log_a(xy) = \log_a x + \log_a y \quad \forall\, x, y \in \mathbb{R}_+ \tag{3.14}$$

da cui segue, in particolare, che $\log_a 1 = 0$. La (3.14) è caratteristica di questa funzione, nel senso precisato dal seguente

■ **Teorema 3.7** - *Sia $f : \mathbb{R}_+ \to \mathbb{R}$ una funzione continua non identicamente nulla. Se vale l'equazione*

$$f(xy) = f(x) + f(y) \quad \forall\, x, y \in \mathbb{R}_+ \tag{3.15}$$

allora $f(x) = \log_a x$ per qualche $a > 0, a \neq 1$.

Dimostrazione - Ponendo $y = 1$ nella (3.15) si trova $f(1) = 0$. Ponendo $y = 1/x$ si trova: $f(1/x) = -f(x)$ sicché possiamo limitarci a considerare uno solo dei due intervalli: $(0,1)$ oppure $(1,+\infty)$; consideriamo, per fissare le idee, l'intervallo $(0,1)$. Applicando ripetutamente la (3.15) si ottengono le regole seguenti:

$$f(x^n) = nf(x) \qquad \forall n \in \mathbb{N},\ x \in (0,1)$$

$$f(x^{m/n}) = \frac{m}{n}\, f(x) \qquad \forall n, m \in \mathbb{N}_+,\ x \in (0,1)$$

e, se $t \in \mathbb{R}$, $t = \lim\limits_{n\to+\infty} r_n$, $r_n \in \mathbb{Q}$, abbiamo

$$\begin{aligned} f(x^t) &= f(x^{\lim r_n}) = f(\lim x^{r_n}) = \lim f(x^{r_n}) = \\ &= \lim r_n f(x) = tf(x) \quad \forall x \in (0,1)\ . \end{aligned}$$

Essendo f non identicamente nulla, esisterà almeno un $b \in (0,1)$ per cui $f(b) \neq 0$. Allora, preso un qualunque $x \in (0,1)$, posto $x = b^{\log_b x}$, abbiamo

$$f(x) = f(b^{\log_b x}) = \log_b x f(b)$$

Posto $a = b^{1/f(b)}$ abbiamo: $f(x) = \log_a x$. □

L'equazione funzionale (3.15) ha una notevole interpretazione probabilistica; essa è legata alla *teoria* classica *dell'informazione*, introdotta da C. Shannon, negli anni quaranta.

I cardini di tale teoria, a livello elementare, sono i seguenti: "informazione" significa sapere se un evento casuale E si è verificato. Si introduce la "quantità di informazione" come funzione della probabilità del verificarsi di E: maggiore è la probabilità di E, minore è la quantità di informazione che si è guadagnata sapendo che si è verificato.

Indichiamo con p la probabilità e con $I(p)$ la quantità di informazione. Circa la funzione $I(p)$ faremo le seguenti ipotesi:

i) $I(p)$ è decrescente (in accordo con quanto sopra detto)

ii) $I(1) = 0$ (ovvero: sapere che si è verificato l'evento certo non aumenta le nostre informazioni)

iii) $\lim_{p \to 0+} I(p) = +\infty$ (ovvero: più l'evento è raro e più grande diventa l'informazione guadagnata nel sapere che si è verificato).

Per identificare $I(p)$ facciamo anche quest'altra ipotesi: siano E ed F eventi indipendenti, cioè il verificarsi dell'uno non influenza il verificarsi dell'altro (*); allora l'informazione guadagnata sapendo che si sono verificati entrambi (cioè si è verificato l'evento $E \cap F$) è pari alla somma delle informazioni guadagnate sapendo che si sono verificati singolarmente. In formule, se p è la probabilità di E e q è la probabilità di F, e dunque pq è la probabilità di $E \cap F$, avremo:

iv) $I(pq) = I(p) + I(q)$.

Dunque sarà (assumendo che $I(p)$ sia continua): $I(p) = \log_a p$ per qualche a. Quale sia questo a dipende dall'unità di misura scelta per la quantità di informazione. Osserviamo comunque che dalla *i*) deve essere $0 < a < 1$.

Una comune unità di misura è il **bit** (*bynary digit*) che corrisponde ad assegnare quantità di informazione uguale ad 1 al verificarsi di uno tra due soli eventi possibili (*acceso–spento*, *vero–falso*, 1 o 0, *magnetizzato–smagnetizzato*) supposti equiprobabili, cioè con probabilità $1/2$ per il singolo evento. In questo caso si avrebbe $1 = I(1/2) = \log_a(1/2)$ da cui $a = 1/2$; si conclude che $I(p) = -\log_2 p$.

3.5 Funzioni iperboliche e loro inverse

Per le ragioni in parte già viste e altre che vedremo nel prossimo capitolo, come *base* per la funzione esponenziale si utilizza il numero e (definito in 4.3.3). Nelle applicazioni della matematica sono importanti certe combinazioni delle funzioni e^x e e^{-x} che ora definiamo:

Sh : *seno iperbolico* $\quad \operatorname{Sh} x := \dfrac{e^x - e^{-x}}{2}$

Ch : *coseno iperbolico* $\quad \operatorname{Ch} x := \dfrac{e^x + e^{-x}}{2}$

Th : *tangente iperbolica* $\quad \operatorname{Th} x := \dfrac{\operatorname{Sh} x}{\operatorname{Ch} x} = \dfrac{e^x - e^{-x}}{e^x + e^{-x}}$.

Esse sono evidentemente definite su tutto $\mathbb{R}$ e continue.

(*) La corretta definizione di eventi indipendenti richiede il concetto di probabilità condizionale, per cui rimandiamo ai testi di Probabilità. Ricordiamo soltanto che, se E ed F sono eventi indipendenti, vale il teorema di moltiplicazione:

$$P(E \cap F) = P(E) \cdot P(F)$$

essendo $P(\cdot)$ la funzione probabilità.

Le seguenti proprietà discendono immediatamente dalla definizione:

i) $\mathrm{Sh}(-x) = -\mathrm{Sh}\,x$ (funzione dispari)
$\mathrm{Ch}(-x) = \mathrm{Ch}\,x$ (funzione pari)
$\mathrm{Th}(-x) = -\mathrm{Th}\,x$ (funzione dispari)

ii) $\mathrm{Sh}(0) = 0 \qquad \mathrm{Ch}(0) = 1 \qquad \mathrm{Th}(0) = 0$
$\mathrm{im}(\mathrm{Sh}) = \mathbb{R} \qquad \mathrm{im}(\mathrm{Ch}) = [1, +\infty) \qquad \mathrm{im}(\mathrm{Th}) = (-1, +1)$
$\mathrm{Sh}\,x \le \frac{1}{2}e^x \le \mathrm{Ch}\,x \qquad \forall x \in \mathbb{R}.$

I grafici di queste funzioni sono illustrati ai n. 16 e 17.

iii) $(\mathrm{Ch}\,x)^2 - (\mathrm{Sh}\,x)^2 = 1$
$\mathrm{Sh}(x+y) = \mathrm{Sh}\,x\,\mathrm{Ch}\,y + \mathrm{Sh}\,y\,\mathrm{Ch}\,x$
$\mathrm{Ch}(x+y) = \mathrm{Ch}\,x\,\mathrm{Ch}\,y + \mathrm{Sh}\,y\,\mathrm{Sh}\,x$
$\mathrm{Th}(x+y) = \dfrac{\mathrm{Th}\,x + \mathrm{Th}\,y}{1 + \mathrm{Th}\,x\,\mathrm{Th}\,y}$
$\mathrm{Sh}(2x) = 2\,\mathrm{Sh}\,x\,\mathrm{Ch}\,x$
$\mathrm{Ch}(2x) = (\mathrm{Ch}\,x)^2 + (\mathrm{Sh}\,x)^2.$

L'aggettivo "iperbolico" dato a queste funzioni deriva dal fatto che esse hanno con l'iperbole una relazione speciale, che ora illustriamo. Se rappresentiamo (v. Fig. 5.6), nel piano (ξ, η), il punto P di coordinate $\xi = \mathrm{Ch}\,t$, $\eta = \mathrm{Sh}\,t$, $t \in \mathbb{R}$, notiamo che, grazie alla prima delle *iii)*, esso si trova sul ramo dell'iperbole di equazione: $\xi^2 - \eta^2 = 1$ contenuto nel semipiano: $\xi > 0$.

Fig. 5.6

Facendo percorrere a t l'asse reale da $-\infty$ a $+\infty$ il punto P percorre questo ramo d'iperbole nel verso indicato in figura. Considerando i triangoli simili POQ e ROV si vede che:

$$QP = \operatorname{Sh} t$$
$$OQ = \operatorname{Ch} t$$
$$VR = \operatorname{Th} t \ .$$

L'allievo confronterà tutte queste proprietà delle funzioni iperboliche con quelle delle funzioni circolari definite nel prossimo paragrafo.

Si vede facilmente che Sh e Th sono monotone strettamente crescenti; perciò esistono le funzioni inverse:

$$\mathbb{R} \to \mathbb{R} \ : \ x \mapsto \operatorname{Sh}^{-1} x$$

$$(-1,+1) \to \mathbb{R} \ : \ x \mapsto \operatorname{Th}^{-1} x$$

e sono continue. Invece Ch, essendo pari, non è invertibile; tuttavia essa è strettamente crescente per $x > 0$ e perciò risulta invertibile la seguente restrizione

$$\operatorname{Ch} \ : [0,+\infty) \to \mathbb{R}$$

che, per abuso di notazione, si indica ancora con Ch. Possiamo definire così la sua funzione inversa

$$\operatorname{Ch}^{-1} \ : [1,+\infty) \to [0,+\infty) \ .$$

I grafici di queste funzioni sono illustrati ai n. 20 e 21.

Una rappresentazione più esplicita di tali funzioni si ottiene per mezzo del logaritmo (in base e). Infatti, preso un qualunque $x \in \mathbb{R}$, esiste un ben determinato y tale che:

$$x = \operatorname{Sh} y = \frac{e^y - e^{-y}}{2}$$

e cioè: $e^y - 2x - e^{-y} = 0$, da cui, moltiplicando per e^y, segue: $e^{2y} - 2xe^y - 1 = 0$. Risolvendo l'equazione quadratica in e^y si ottiene $e^y = x + \sqrt{x^2+1}$ (la soluzione: $x - \sqrt{x^2+1}$ va scartata, perché negativa) e infine, passando ai logaritmi: $y = \log(x + \sqrt{1+x^2})$, per ogni $x \in \mathbb{R}$.

Perciò si ha:

$$\operatorname{Sh}^{-1} x = \log\left(x + \sqrt{x^2+1}\right) \qquad x \in \mathbb{R} \ . \tag{3.16}$$

Analogamente si trova:

$$\operatorname{Ch}^{-1} x = \log\left(x + \sqrt{x^2-1}\right) \qquad x \in [1,+\infty) \tag{3.17}$$

$$\operatorname{Th}^{-1} x = \frac{1}{2}\log\frac{1+x}{1-x} \qquad x \in (-1,+1) \ . \tag{3.18}$$

Per ragioni che saranno chiare più avanti, le funzioni $\mathrm{Sh}^{-1}, \mathrm{Ch}^{-1}, \mathrm{Th}^{-1}$ vengono anche indicate con i simboli

$$\mathrm{Sett\,Sh} \quad , \quad \mathrm{Sett\,Ch} \quad , \quad \mathrm{Sett\,Th}$$

rispettivamente (si legge: settore senoiperbolico, e simili).

3.6 Funzioni circolari (o trigonometriche) e loro inverse

Assieme alle funzioni esponenziali (e ai logaritmi) in Matematica hanno pari importanza le funzioni circolari o trigonometriche, che lo studente già conosce, ma che qui vogliamo richiamare per metterne in risalto alcune proprietà. L'importanza di queste due classi di funzioni deriva dal fatto che esse sono in un certo senso i prototipi utili à descrivere due corrispondenti gruppi di fenomeni frequentissimi in natura: i fenomeni di decadimento (o al contrario, di crescita) e i fenomeni periodici. Esempi di fenomeni del primo tipo sono: il decadimento radioattivo, il processo di raffreddamento di un corpo, il diffondersi di un'infezione, o il moltiplicarsi di una colonia di batteri. Esempi di fenomeni del secondo tipo sono: il moto dei pianeti, il moto del pendolo, certi andamenti di malattie influenzali di carattere stagionale, etc. Ora accade, e l'allievo lo osserverà nel corso dei suoi studi, che spesso le funzioni esponenziali servono a descrivere i fenomeni del primo tipo (naturalmente, nei casi più semplici) mentre quelle trigonometriche sono utili a descrivere i fenomeni del secondo tipo.

Definizione 3.3 - *Una funzione reale o complessa f di variabile reale o complessa è detta* **periodica**, *con periodo $p \neq 0$, se per ogni $x \in \mathrm{dom}(f)$, anche $x + p \in \mathrm{dom}(f)$ e risulta*

$$f(x+p) = f(x)\,. \tag{3.19}$$

Occupiamoci ora di funzioni periodiche di variabile reale. Se p è un periodo, anche $-p, 2p, np$ (con $n \in \mathbb{Z}$) sono periodi. Consideriamo i periodi positivi: o l'estremo inferiore di questi periodi è 0 oppure è $p_0 > 0$: diremo allora che la funzione è periodica di periodo (minimo) p_0. Il reciproco del periodo vien detto *frequenza*. Si osserverà che l'insieme delle funzioni periodiche con lo stesso periodo forma uno spazio lineare.

Veniamo ora alle funzioni trigonometriche seno e coseno, che, come è noto, sono funzioni periodiche con periodo 2π.

Nel piano euclideo $\mathbb{R}^2$ fissiamo un'origine O e consideriamo la circonferenza S con centro in O e raggio unitario. Fissiamo un verso di percorrenza su S (generalmente quello antiorario) e definiamo *angolo orientato* una *coppia ordinata* di raggi (r, r') uscenti da O (analogamente, sulla retta orientata, si definisce *segmento orientato* una coppia ordinata di punti).

La misura di un angolo si fa adottando una opportuna unità di misura. Consideriamo l'angolo orientato definito dai due raggi r, r' che stanno sulla stessa retta e di verso opposto. Si adotta la misura in *radianti* assegnando a quest'angolo misura π; si adotta la misura in *gradi (sessagesimali)* ponendo la misura di quest'angolo uguale a 180. Adottando, come sempre faremo, salvo contrario avviso, la misura in radianti, lo stesso numero esprime la misura dell'angolo e dell'arco (orientato) della circonferenza di raggio unitario che sottende quest'angolo.

Fissato un punto su S (per esempio il punto A in figura 5.7 di coordinate cartesiane $(1,0)$) si misurano gli angoli a partire dal raggio OA; una misura $t > 0$ significa che, a partire da A, si è percorso un arco di ampiezza t nel verso antiorario, se $t < 0$ si è percorso un arco di ampiezza $|t|$ nel verso orario.

In tal modo, ad ogni $t \in \mathbb{R}$ corrisponde un preciso punto $P \in S$: quel punto per cui l'angolo (OA, OP) ha misura t.

Indichiamo con $\Phi : \mathbb{R} \to S$ tale applicazione. Con riferimento alla figura 5.7 abbiamo:

$$\Phi(0) = A,\ \ \Phi\left(\frac{\pi}{2}\right) = B,\ \ \Phi(\pi) = C,\ \ \Phi\left(\frac{3}{2}\pi\right) = D$$

$$\Phi(2\pi) = A,\ \ \Phi\left(-\frac{\pi}{2}\right) = D,\ \ \Phi(-\pi) = C...$$

Si vede così che tale applicazione *non* è *iniettiva* e inoltre essa risulta *periodica* di periodo 2π:

$$\Phi(t + 2\pi) = \Phi(t)\ .$$

Se indichiamo con (ξ, η) le coordinate cartesiane del punto $P \in S$, sarà ovviamente: $\xi^2 + \eta^2 = 1$. L'applicazione $\Phi : \mathbb{R} \to S$ può allora essere vista come una coppia di applicazioni:

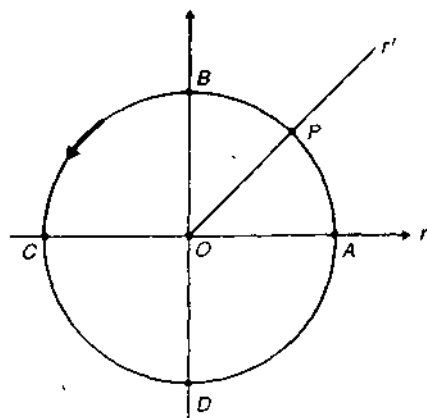



Fig. 5.7

$$\Phi = (\Phi_1, \Phi_2)$$
$$\Phi_1 : \mathbb{R} \to \mathbb{R} \quad , \; \Phi_1(t) = \xi$$
$$\Phi_2 : \mathbb{R} \to \mathbb{R} \quad , \; \Phi_2(t) = \eta$$
$$\Phi_1(t)^2 + \Phi_2(t)^2 = 1 \; .$$

Sia Φ_1 che Φ_2 sono periodiche di periodo 2π. Le funzioni Φ_1 e Φ_2 vengono indicate con i simboli ben noti:

$$\Phi_1 := \cos \qquad , \qquad \Phi_2 := \sin \; .$$

Dalla definizione seguono poi le proprietà che qui appresso elenchiamo. È utile confrontare queste proprietà con quelle delle funzioni iperboliche elencate in 3.5.

$$\mathrm{im}(\cos) = [-1, +1]$$

$$\mathrm{im}(\sin) = [-1, +1]$$

$$\sin(-x) = -\sin x \qquad \text{(funzione dispari)}$$

$$\cos(-x) = \quad \cos x \qquad \text{(funzione pari)}$$

$$(\cos x)^2 + (\sin x)^2 = 1$$

$$\left.\begin{array}{l} \cos(x+y) = \cos x \cos y - \sin x \sin y \\ \sin(x+y) = \sin x \cos y + \cos x \sin y \end{array}\right. \qquad \text{(formule di addizione)}$$

$$\left.\begin{array}{l} \sin(2x) = 2 \sin x \cos x \\ \cos(2x) = (\cos x)^2 - (\sin x)^2 \end{array}\right. \qquad \text{(formule di duplicazione)}$$

$$\left.\begin{array}{l} \cos\left(\frac{\pi}{2} - x\right) = \sin x \\ \sin\left(\frac{\pi}{2} - x\right) = \cos x \end{array}\right. \qquad \text{(archi complementari)}$$

$$\left.\begin{array}{l} \sin x - \sin y = 2 \sin \frac{x-y}{2} \cos \frac{x+y}{2} \\ \cos x - \cos y = -2 \sin \frac{x-y}{2} \sin \frac{x+y}{2} \end{array}\right. \qquad \text{(formule di prostaferesi)}$$

$$0 < \sin x < x \;\; \text{per} \;\; 0 < x < \frac{\pi}{2} \; .$$

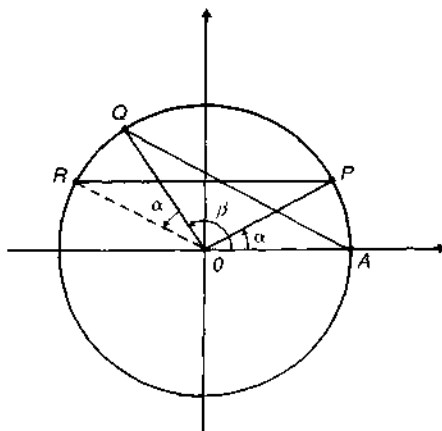



Fig. 5.8

A titolo di esempio, dimostriamo la formula di addizione per il coseno. Con riferimento alla figura 5.8 P ha coordinate $(\cos\alpha, \sin\alpha)$, Q $(\cos\beta, \sin\beta)$ e R $(\cos(\alpha+\beta), \sin(\alpha+\beta))$; inoltre le corde $\overline{RP}$ e $\overline{AQ}$ hanno uguale misura poiché sottendono angoli al centro di uguale misura. Abbiamo perciò

$$\begin{aligned}\overline{RP}^2 &= (\cos(\alpha+\beta) - \cos\alpha)^2 + (\sin(\alpha+\beta) - \sin\alpha)^2 = \\ &= \overline{AQ}^2 = (1-\cos\beta)^2 + (\sin\beta)^2\end{aligned}$$

da cui si ottiene

$$\cos\alpha\cos(\alpha+\beta) + \sin\alpha\sin(\alpha+\beta) = \cos\beta\ .$$

Ponendo $\alpha = -x$ e $\beta = y + x$ otteniamo il risultato voluto.

Abbiamo già dimostrato che le funzioni sin e cos sono continue su $\mathbb{R}$. I loro grafici sono illustrati ai n. 27 e 28.

Si definisce poi la funzione

$$\mathrm{tg} : \left\{x \in \mathbb{R} : x \neq \frac{\pi}{2} + k\pi, k \in \mathbb{Z}\right\} \to \mathbb{R}$$

$$\mathrm{tg}\, x := \frac{\sin x}{\cos x}\ .$$

Essa risulta continua su tutto l'insieme di definizione e periodica di periodo π. Inoltre valgono le proprietà:

$$\mathrm{im}(\mathrm{tg}) = (-\infty, +\infty)$$

$$\mathrm{tg}(-x) = -\,\mathrm{tg}\,x \quad \text{(funzione dispari)}$$

$$\mathrm{tg}(x+y) = \frac{\mathrm{tg}\,x + \mathrm{tg}\,y}{1 - \mathrm{tg}\,x\,\mathrm{tg}\,y} \quad \text{(formula di addizione).}$$

Si definiscono anche le funzioni:

$$\mathrm{cotg}\,x := \frac{\cos x}{\sin x} \quad , \quad \sec x := \frac{1}{\cos x} \quad , \quad \mathrm{cosec}\,x := \frac{1}{\sin x} \; .$$

(che prendono il nome di: cotangente, secante, cosecante rispettivamente).

Le figure 5.9 a, b, c, d illustrano geometricamente i rapporti tra le tre funzioni trigonometriche principali.

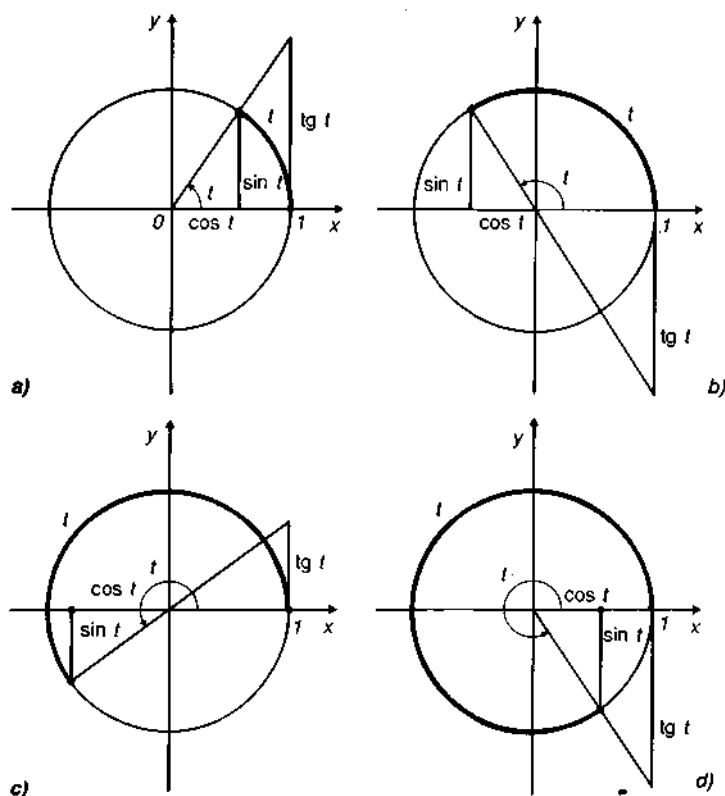



Fig. 5.9

Tutte le funzioni considerate in questo paragrafo, essendo periodiche, *non* sono *invertibili.* Possiamo tuttavia parlare di *relazioni inverse* delle funzioni trigonometriche; i grafici di queste relazioni si ottengono da quelli delle corrispondenti funzioni per simmetria rispetto alla bisettrice del 1° e 3° quadrante; essi sono riportati ai n. 29, 30, 32.

È chiaro che non si tratta di funzioni, ma di *multifunzioni.* Per esempio, il dominio di definizione della relazione inversa di sin è $\mathrm{im}(\sin)$, cioè $[-1,+1]$: ad ogni $x \in [-1,+1]$ la relazione fa corrispondere un'infinità di valori:

$$0 \mapsto \{\cdots - 2\pi, -\pi, 0, \pi, 2\pi, \cdots\}$$

$$1 \mapsto \{\frac{\pi}{2} + 2k\pi\}_{k \in \mathbb{Z}} .$$

Tuttavia, opportune restrizioni delle funzioni trigonometriche sono iniettive; per esempio, le funzioni (che, per abuso di notazione, indichiamo con gli stessi simboli già utilizzati)

$$\sin : \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \to \mathbb{R}$$

$$\cos : [0, \pi] \to \mathbb{R}$$

$$\mathrm{tg} : \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \to \mathbb{R}$$

$$\mathrm{cotg} : (0, \pi) \to \mathbb{R}$$

sono iniettive e pertanto invertibili; le *funzioni inverse* si indicano rispettivamente con i simboli

$\sin^{-1}$	o anche	$\arcsin$:	$[-1,+1] \to \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$
$\cos^{-1}$	o anche	$\arccos$:	$[-1,+1] \to [0, \pi]$
tg^{-1}	o anche	arctg :	$\mathbb{R} \to \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$
cotg^{-1}	o anche	$\mathrm{arccotg}$:	$\mathbb{R} \to (0, \pi)$

(arcsin si legge: arcoseno e similmente per gli altri simboli). I loro grafici sono indicati con tratto pieno nei grafici delle relazioni inverse.

Bisogna osservare che talvolta i simboli sopra riportati sono utilizzati per indicare le *relazioni* inverse mentre, per indicare le *funzioni*, si dice: *determinazione principale di* arcsin, etc.

3.7 Esponenziale complesso

Riprendiamo in esame l'applicazione $\Phi : \mathbb{R} \to S$ definita nel precedente paragrafo. Identifichiamo $\mathbb{R}^2$ con $\mathbb{C}$; i punti di S sono allora numeri complessi di modulo 1. Osserviamo subito che:

$$z_1, z_2 \in S \Rightarrow z_1 z_2 \in S$$

$$z \in S \Rightarrow \frac{1}{z} = \overline{z} \in S$$

$$1 \in S \quad .$$

Perciò $S \subset \mathbb{C}$ forma un gruppo commutativo rispetto al prodotto: $(S, \cdot)$; inoltre vale l'uguaglianza, per ogni $t_1, t_2 \in \mathbb{R}$,

$$\Phi(t_1 + t_2) = \Phi(t_1)\Phi(t_2) \, .$$

L'applicazione Φ manda il gruppo commutativo $(\mathbb{R}, +)$ nel gruppo commutativo $(S, \cdot)$. Questa osservazione rivela l'esistenza di una parentela formale tra la funzione esponenziale e le funzioni trigonometriche; ma indica anche che questa parentela viene messa in luce nel campo complesso.

Sviluppiamo questa osservazione estendendo l'applicazione Φ a tutto $\mathbb{C}$; indichiamo con Φ_e questa estensione. Se $z = (x, y) \in \mathbb{C}$, poniamo per definizione:

$$\Phi_e : \mathbb{C} \to \mathbb{C}$$

$$\Phi_e(z) := (e^x \cos y, e^x \sin y) \, .$$

Perciò la restrizione di Φ_e data da: $(0, y) \mapsto (\cos y, \sin y)$ si può identificare con la precedente applicazione Φ, mentre la restrizione: $(x, 0) \mapsto (e^x, 0)$ si può identificare con la funzione esponenziale reale (con base e): exp. Inoltre, come è facile verificare, vale l'uguaglianza, per ogni $z_1, z_2 \in \mathbb{C}$:

$$\Phi_e(z_1 + z_2) = \Phi_e(z_1)\Phi_e(z_2) \, .$$

Per analogia con la nomenclatura e la notazione già usata nel caso reale, l'applicazione Φ_e viene chiamata *esponenziale* (*complesso*) e si pone:

$$\Phi_e(z) := \exp z \quad \text{o anche} \quad e^z \, .$$

Si pone cioè:

$$e^z := e^x \cos y + i e^x \sin y \, .$$

In particolare, se $z = x$ è reale ($y = 0$) si ritrova e^x; se $z = iy$ è immaginario puro ($x = 0$) risulta

$$e^{iy} = \cos y + i \sin y \quad \text{(formula di Eulero).} \tag{3.20}$$

La formula di Eulero consente di rappresentare un numero complesso z in una forma compatta, detta *forma esponenziale*, scrivendo $z = re^{i\theta}$ dove $r = |z|$ è il modulo e θ un argomento di z.

I numeri complessi possono dunque essere rappresentati in:

forma cartesiana: $z = x + iy \quad x, y \in \mathbb{R}$

forma trigonometrica: $z = r(\cos\theta + i\sin\theta) \quad r \in \overline{\mathbb{R}}_+, \theta \in \mathbb{R}$

forma esponenziale: $z = re^{i\theta}$.

In particolare abbiamo, $\forall\, k \in \mathbb{Z}$:

$$1 = e^0 = e^{2\pi i} = e^{2k\pi i}$$
$$-1 = e^{\pi i} = e^{-\pi i} = e^{(2k+1)\pi i} \ .$$

(È curioso osservare che l'ultima formula può essere scritta nella forma $e^{i\pi} + 1 = 0$ che fa intervenire i più importanti numeri della matematica: $0, 1, \pi, e, i$).

$$i = e^{(\pi/2)i} = e^{i\pi(1/2+2k)}$$

$$-i = e^{-(\pi/2)i} = e^{i\pi(-1/2+2k)} \ .$$

In generale abbiamo

$$e^{z+2k\pi i} = e^z \qquad \forall k \in \mathbb{Z} \ . \tag{3.21}$$

Perciò l'esponenziale complesso è funzione periodica, con periodo $2\pi i$. Esso non è dunque una funzione iniettiva. Mostriamo ora che:

$$\operatorname{im}(\exp) = \mathbb{C} \setminus \{0\} \tag{3.22}$$

(e pertanto, non essendo mai nulla, $\exp z$ non è una funzione razionale).

Infatti

$$e^z = 0 \;\Leftrightarrow\; \begin{cases} e^x \cos y = 0 \\ e^x \sin y = 0 \end{cases} \;\Rightarrow\; \begin{cases} \cos y = 0 \\ \sin y = 0 \end{cases}$$

perciò $0 \notin \operatorname{im}(\exp)$. Inoltre, se $w \in \mathbb{C} \setminus \{0\}$, scrivendo $w = r(\cos\theta + i\sin\theta)$, basta porre $x = \log r$, $y = \theta$ e risulta $w = e^z$, cioè $w \in \operatorname{im}(\exp)$.

Infine la funzione $\exp : \mathbb{C} \to \mathbb{C}$ è continua perché le sue componenti (parte reale e parte immaginaria) sono continue.

A questo punto è facile estendere anche le funzioni sin e cos al campo complesso. Infatti, dalla formula di Eulero (3.20) combinata con l'analoga

$$e^{-iy} = \cos y - i\sin y \tag{3.23}$$

(che si ottiene dalla (3.20) cambiando y in $-y$) per somma e sottrazione si ottiene:

$$\sin y = \frac{e^{iy} - e^{-iy}}{2i} \tag{3.24}$$

$$\cos y = \frac{e^{iy} + e^{-iy}}{2} \,. \tag{3.25}$$

Poniamo allora, per definizione, per ogni $z \in \mathbb{C}$:

$$\sin z := \frac{e^{iz} - e^{-iz}}{2i} \tag{3.26}$$

$$\cos z := \frac{e^{iz} + e^{-iz}}{2} \tag{3.27}$$

$$\operatorname{tg} z := \frac{\sin z}{\cos z} = \frac{1}{i}\frac{e^{iz} - e^{-iz}}{e^{iz} + e^{-iz}} \,. \tag{3.28}$$

È facile verificare che tutte le proprietà formali (periodicità, simmetrie, formule di addizione, di duplicazione, di prostaferesi, etc.) restano valide.

Estendendo infine anche le funzioni iperboliche al campo complesso, ponendo

$$\operatorname{Sh} z := \frac{e^{z} - e^{-z}}{2} \tag{3.29}$$

$$\operatorname{Ch} z := \frac{e^{z} + e^{-z}}{2} \tag{3.30}$$

$$\operatorname{Th} z := \frac{\operatorname{Sh} z}{\operatorname{Ch} z} \tag{3.31}$$

si trova il legame tra le funzioni circolari e le funzioni iperboliche:

$$\cos z = \operatorname{Ch}(iz) \tag{3.32}$$

$$\sin z = -i \operatorname{Sh}(iz) \,. \tag{3.33}$$

Si osservi che tutte le funzioni sopra definite, che sono continue in $\mathbb{C}$, non ammettono limite per $z \to \infty$ (lo studente si accerti di questo fatto calcolando il limite per $z \to \infty$ in varie direzioni).

3.8 Logaritmo complesso. Operazione di elevamento a potenza nel campo complesso

Come abbiamo già osservato, la funzione esponenziale essendo, nel campo complesso, periodica, non è invertibile. La *relazione* inversa di exp è detta *logaritmo complesso* e viene indicata con $\log_e$ o anche semplicemente con log: in tal caso

però deve risultare chiaro dal contesto che si tratta del logaritmo complesso e non del logaritmo reale. In questo paragrafo useremo la notazione esplicita $\log_c$.

Dalla definizione segue allora che, se $z \in \mathbb{C}$ e $w \in \mathbb{C}\setminus\{0\}$:

$$e^z = w \Leftrightarrow z = \log_c w \ .$$

Posto $z = x + iy$ e $w = \rho e^{i\varphi}$ l'equazione $e^z = w$ si scrive:

$$e^x e^{iy} = \rho e^{i\varphi} \ .$$

Ricordando che due numeri complessi sono uguali se hanno uguali i moduli e gli argomenti differiscono per multipli di 2π, dalla precedente uguaglianza deduciamo:

$$e^x = \rho \quad \text{cioè} \quad x = \log\rho$$

$$y = \varphi + 2k\pi \quad \text{con} \quad k \in \mathbb{Z} \ .$$

Abbiamo perciò, per ogni numero complesso $w = \rho e^{i\varphi}$, $w \neq 0$:

$$\log_c w = \log\rho + i(\varphi + 2k\pi) \ . \tag{3.34}$$

Esistono perciò, $\forall w \in \mathbb{C}\setminus\{0\}$, infiniti logaritmi di w; essi hanno tutti la stessa parte reale (il logaritmo reale del modulo di w: $\log|w|$) e parti immaginarie che si ottengono da una di esse ($= \arg w$) aggiungendo multipli di 2π.

***Esempio* 3.11 -**

$$\log_c(-1) = \log|-1| + i(\pi + 2k\pi) = \pi i(2k+1) \quad , \ k \in \mathbb{Z}$$

$$\log_c 1 = \log 1 + i(0 + 2k\pi) = 2k\pi i \ .$$

Si noti che, mentre tutti i logaritmi complessi di -1 hanno parte immaginaria diversa da zero, tra gli infiniti logaritmi complessi di 1, uno è reale: il logaritmo reale di 1. In generale, se w è reale positivo, il suo logaritmo reale è *uno* degli infiniti logaritmi complessi di w: l'unico reale.

Le proprietà formali dei logaritmi restano valide anche nel campo complesso. In particolare, $\forall z_1, z_2 \in \mathbb{C}\setminus\{0\}$, vale:

$$\log_c(z_1 z_2) = \log_c z_1 + \log_c z_2 \tag{3.35}$$

Bisogna notare però che la (3.35), essendo un'uguaglianza tra *multifunzioni*, è vera nel *senso insiemistico*, e cioè: se $\log_c z_1$ e $\log_c z_2$ denotano l'insieme dei

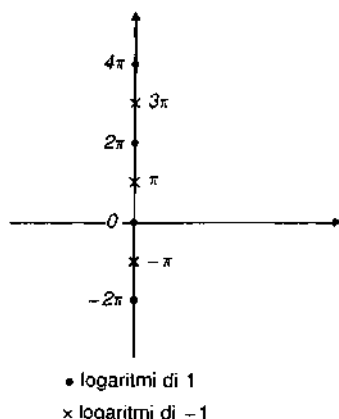



Fig. 5.10

logaritmi complessi di z_1 e z_2 rispettivamente, allora l'insieme formato sommando tutti questi elementi (un addendo nel primo insieme, l'altro nell'altro) coincide con l'insieme dei logaritmi complessi di $z_1 z_2$.

Naturalmente, come abbiamo già fatto in precedenza, possiamo considerare una opportuna restrizione della funzione esponenziale, per esempio:

$$\exp : \{z \in \mathbb{C} : -\pi < \operatorname{Im} z \leq \pi\} \to \mathbb{C} \ .$$

Questa allora è iniettiva e si può parlare di *funzione inversa*; essa vien detta *determinazione principale* del logaritmo complesso. È una funzione continua in $\mathbb{C} \setminus \{0\}$; non ammette limite per $z \to 0$ né per $z \to \infty$.

Negli esercizi 10 e 11 si ricaveranno le formule che permettono di esprimere le funzioni trigonometriche inverse per mezzo del logaritmo complesso.

La definizione della funzione exp e della relazione inversa $\log_c$ ci permette ora di estendere l'operazione di elevamento a potenza, definita in 2.3.2 solo per base reale positiva ed esponente reale, al campo complesso.

Definizione 3.4 - *Dato* $t \in \mathbb{C} \setminus \{0\}$, $z \in \mathbb{C}$, *poniamo*

$$t^z := e^{z \log_c t} \ . \tag{3.36}$$

La definizione è evidentemente giustificata dal fatto che

$$\exp(\log_c w) = w, \ \forall w \in \mathbb{C} \setminus \{0\} \quad \text{e} \quad \log_c t^z = z \log_c t \ .$$

È evidente che l'operazione non è più, in generale, univocamente definita; accade, come già per l'estrazione di radice n–esima nel campo complesso, che questa operazione abbia un insieme di risultati, magari infinito. Tuttavia, come si verifica subito, le proprietà formali dell'operazione, in particolare la proprietà fondamentale

$$t^{z_1 + z_2} = t^{z_1} t^{z_2} \tag{3.37}$$

restano valide (pur di intenderle, come abbiamo già spiegato, in senso insiemistico).

Esempi

3.12. Sia a reale > 0; calcoliamo a^3 in senso complesso. Si ha, per $k \in \mathbb{Z}$

$$a^3 = e^{3 \log_c a} = e^{3 \log a + 3 \cdot 2k\pi i} = a^3 \text{ (in senso reale).}$$

In questo caso dunque l'operazione è univocamente definita e coincide con l'analoga operazione già definita sul campo reale.

3.13. $i^{\frac{1}{4}} = e^{\frac{1}{4} \log_c i} = e^{\frac{1}{4}(\log |i| + i(\pi/2 + 2k\pi))}$ $(k \in \mathbb{Z}) =$
$= e^{i\pi/8} e^{i\frac{\pi}{2}k} = \{e^{i\pi/8}, i\, e^{i\pi/8}, -e^{i\pi/8}, -i\, e^{i\pi/8}\}.$

In questo caso l'operazione coincide con l'estrazione di radice quarta, già definita in 2.4.3.

3.14. $(-1)^{\sqrt{2}} = e^{\sqrt{2} \log_c (-1)} = e^{\pi i \sqrt{2}(2k+1)}, k \in \mathbb{Z}$.

L'operazione, che non ha senso nel campo reale, ha una infinità di risultati, tutti complessi.

3.15. $i^i = e^{i \log_c i} = e^{i(\log |i| + i(\pi/2 + 2k\pi))}(k \in \mathbb{Z}) =$
$\doteq e^{-\pi(1/2 + 2k)}.$

L'operazione ha una infinità di risultati tutti reali! (ma l'operazione non ha significato nel campo reale!).

Osservazione - Si potrebbe notare che può nascere ambiguità sulla notazione e^z, una volta intesa come immagine di z data dalla funzione exp, ben definita, oppure intesa come operazione di elevamento a potenza con base e, definita dalla (3.36); in tal caso sarebbe:

$$e^z = \exp(z \log_c e) = \exp\{z(1 + 2k\pi i)\}, \ (k \in \mathbb{Z})$$
$$= \exp(z) \cdot \exp(2k\pi i z) \ .$$

Si conviene che, ogniqualvolta si scrive e^z, si intenda l'immagine (univocamente definita!) del numero z data dalla funzione esponenziale.

Si chiamano *funzioni elementari* (da $\mathbb{C}$ in $\mathbb{C}$) dell'Analisi le seguenti:

1) Funzioni razionali (intere e fratte)

2) Funzioni algebriche (esplicite ed implicite)

3) La funzione esponenziale

4) La funzione logaritmo

5) Tutte quelle funzioni che si ottengono combinando o componendo un numero finito di volte funzioni del tipo 1)...4).

Perciò sono funzioni elementari anche le funzioni trigonometriche e le iperboliche e le loro inverse. Le funzioni elementari che non sono razionali o algebriche si dicono *trascendenti*. Tutte le funzioni elementari sono continue. Esse formano uno spazio lineare (sul campo complesso); esse formano anche un corpo commutativo (campo) rispetto alle solite operazioni di somma e prodotto di funzioni.

Vedremo nei prossimi capitoli che l'insieme delle funzioni elementari possiede speciali proprietà rispetto alle operazioni di derivazione e integrazione.

Esercizi

5. Dimostrare che, se $\mathcal{P} \subset \mathbb{R}_+$ è un insieme di periodi per la funzione f ed f è continua, anche $\inf \mathcal{P}$ è un periodo; se $\inf \mathcal{P} = 0$, $f =$ costante.

6. Dimostrare che l'insieme delle funzioni periodiche $f : \mathbb{R} \to \mathbb{R}$ di periodo razionale forma uno spazio lineare. L'insieme di *tutte* le funzioni periodiche da $\mathbb{R}$ in $\mathbb{R}$ è uno spazio lineare?

7. Trovare gli zeri di $\cos z, \sin z, \mathrm{Ch}\, z, \mathrm{Sh}\, z$; dedurre che queste funzioni non sono razionali.

8. Trovare tutte le soluzioni dell'equazione: $\sin z = 2$.

9. Mostrare che le funzioni $\sin z, \cos z, \mathrm{Sh}\, z, \mathrm{Ch}\, z$ non sono limitate e non ammettono limite per $z \to \infty$.

10. Risolvere le equazioni: $\sin z = t$, $\cos z = t$ per ogni $t \in \mathbb{C}$ e dedurre le seguenti espressioni per le relazioni $\arcsin x$ e $\arccos x$:

$$\arcsin x = -i \log_e (ix \pm \sqrt{1 - x^2})$$

$$\arccos x = -i \log_e (x \pm \sqrt{x^2 - 1}) \, .$$

11. Ricavare la seguente formula per l'arcotangente:

$$\mathrm{arctg}\, x = \frac{1}{2i} \log_e \frac{1 + ix}{1 - ix} \, .$$

12. Calcolare tutte le determinazioni di:

$$a)\ \log_e(1 + i) \quad b)\ 1^i \quad c)\ \log_e \frac{1 + i}{1 - i} \, .$$

Calcolo differenziale 1
Funzioni reali di variabile reale 6

Il problema geometrico della determinazione della tangente ad una curva piana condusse verso la fine del 17° secolo ai concetti di derivata e di differenziale. Furono Leibniz e Newton, principalmente, a sviluppare il calcolo differenziale che ne deriva.

Le idee di Leibniz (basate sui concetti di infinitesimo ed infinito attuali) e quelle di Newton (basate sul concetto di "flussione") pur fecondissime sul piano dell'intuizione, rivelavano gravi carenze su quello del rigore; per tale ragione il calcolo differenziale si sviluppa oggi lungo linee completamente diverse, basandosi sulla nozione di limite introdotta pionieristicamente da D'Alembert e poi ripresa sistematicamente da Cauchy agli inizi del secolo scorso per arrivare alla versione definitiva di Weierstrass intorno al 1870. Sul piano del rigore la teoria è ora ineccepibile; si paga un po' su quello dell'immediatezza e dell'intuizione.

Presenteremo tuttavia, nell'appendice alla fine del volume, alcuni elementi di Analisi non standard, disciplina fondata da Abraham Robinson intorno al 1960, che, sfruttando recenti risultati di logica matematica (la teoria dei modelli), permette di ritornare con pieno rigore alle idee ed ai metodi di Leibniz e Newton.

Svilupperemo il calcolo differenziale in questo e nel prossimo capitolo: in questo tratteremo le funzioni reali di una variabile reale, nel Cap. 7 le funzioni di più variabili (da $\mathbb{R}^n$ in $\mathbb{R}$ e da $\mathbb{R}^n$ in $\mathbb{R}^m$) e le funzioni complesse (da $\mathbb{C}$ in $\mathbb{C}$).

Nella sez. 1 si introducono i concetti fondamentali di derivata e differenziale, e si ricavano le regole di calcolo. Nella sez. 2 presentiamo gli sviluppi del calcolo differenziale: i teoremi principali e alcune prime conseguenze. Altre applicazioni, più specifiche e più estese (come le proprietà delle funzioni convesse o applicazioni al calcolo numerico) vengono presentate nella sez. 3.

1. DERIVATA E DIFFERENZIALE

1.1 Definizione di derivata. Derivata destra, derivata sinistra. Derivate successive

Sia data $f:(a,b) \to \mathbb{R}$ e sia $x \in (a,b)$.

Consideriamo il quoziente, detto **rapporto incrementale**

$$\frac{f(x+h)-f(x)}{h}$$

con $h \neq 0$ e $|h|$ abbastanza piccolo, in modo che $x+h \in (a,b)$.

Il significato del rapporto incrementale è quello di *velocità media* o *tasso medio di variazione di* f *rispetto ad* x.

Definizione 1.1 - *Sia* $f\colon(a,b) \to \mathbb{R}$*; si definisce* **derivata** *di* f *nel punto* x *il numero*

$$f'(x) := \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} \tag{1.1}$$

purché il limite esista finito.

Se $f'(x)$ è definito, diremo che f è *derivabile* in x.

Se $f'(x)$ è definito per ogni $x \in E \subseteq (a,b)$, diremo che f è *derivabile* in E. In questo caso risulta definita la funzione

$$f'\colon x \mapsto f'(x)\colon E \to \mathbb{R}$$

che si chiama **derivata** o **derivata prima** di f.

Altre notazioni per la funzione derivata sono Df, df/dx (notazione di Leibniz), $\dot{f}$ (notazione di Newton).

Il significato di f' è quello di *velocità* o *tasso di variazione istantaneo di* f *rispetto ad* x. Ecco alcune interpretazioni concrete con notazioni più appropriate per le variabili.

variabile	interpretazione	funzione	interpretazione	derivata	interpretazione
t	tempo	$s(t)$	spazio percorso al tempo t	$\frac{ds}{dt}$	velocità
t	tempo	$\theta(t)$	angolo coperto al tempo t	$\frac{d\theta}{dt}$	velocità angolare
t	tempo	$v(t)$	velocità	$\frac{dv}{dt}$	accelerazione
t	tempo	$q(t)$	quantità di carica al tempo t	$\frac{dq}{dt}$	intensità di corrente
V	volume	$m(V)$	massa contenuta nel volume V	$\frac{dm}{dV}$	densità (di massa)
p	prodotto	$k(p)$	costo di produzione	$\frac{dk}{dp}$	costo marginale
x	ascissa (rettilinea)	$f(x)$	quota raggiunta al punto x	$\frac{df}{dx}$	pendenza in x

L'ultimo degli esempi mostra il significato geometrico della derivata che permette di risolvere il problema delle tangenti ad una curva piana e che qui esaminiamo in dettaglio.

Consideriamo il grafico di una funzione, come mostrato in figura 6.1.

Dalla trigonometria elementare deduciamo che

$$\frac{f(x_0+h)-f(x_0)}{h} = \operatorname{tg}\omega_h = \text{ coefficiente angolare o pendenza di } r_h .$$

Quando $h \to 0$, P_0 è fisso e Q_h si muove verso P_0 mantenendosi sul grafico di f e la retta r_h cambia direzione.

Sia f derivabile in x_0. Allora $\operatorname{tg}\omega_h \to f'(x_0)$ per $h \to 0$ e cioè le rette r_h definiscono una "*retta limite*" r, con pendenza $f'(x_0)$ che si chiama *retta tangente* al grafico di f nel punto P_0. In riferimento alla figura abbiamo dunque

$$f'(x_0) = \operatorname{tg}\omega \ . \tag{1.2}$$

L'*equazione della retta tangente* nel punto P_0 è:

$$y = f(x_0) + f'(x_0)(x - x_0) \ . \tag{1.3}$$

Esempi

1.1. Dall'interpretazione geometrica della derivata abbiamo:
se $f(x) = k$ costante allora $f'(x) = 0 \quad \forall x \in \mathbb{R}$;
se $f(x) = x$ allora $f'(x) = 1 \quad \forall x \in \mathbb{R}$.

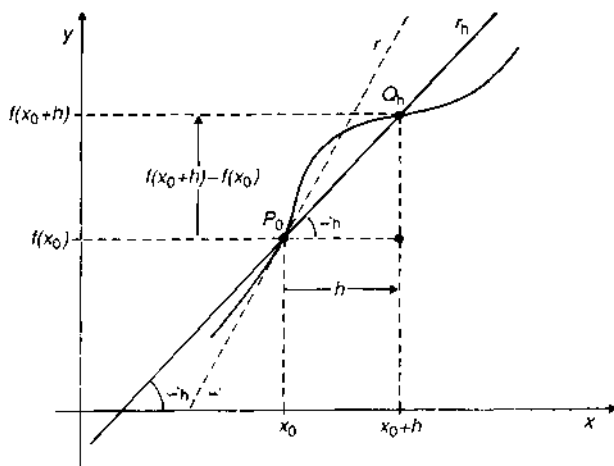



Fig. 6.1

1.2. Sia $f(x) = x^2$. Allora

$$f(x+h) - f(x) = (x+h)^2 - x^2 = 2hx + h^2 ;$$

dunque

$$\frac{f(x+h) - f(x)}{h} = 2x + h \to 2x \text{ per } h \to 0 .$$

Si conclude che f è derivabile in $\mathbb{R}$ e $f'(x) = 2x$.

1.3. Sia $n \in \mathbb{N}$, $n \geq 1$ e $f(x) = x^n$. Allora, dalla formula di Newton,

$$f(x+h) - f(x) = (x+h)^n - x^n = nx^{n-1}h + \binom{n}{2}x^{n-2}h^2 + \cdots + h^n ;$$

dunque

$$\frac{f(x+h) - f(x)}{h} = nx^{n-1} + \binom{n}{2}x^{n-2}h + \cdots + h^{n-1} \longrightarrow nx^{n-1} \text{ per } h \to 0$$

Si conclude che f è derivabile in $\mathbb{R}$ e $f'(x) = nx^{n-1}$.

1.4. Calcoliamo l'equazione della retta tangente alla cubica $f(x) = x^3$ nel punto del suo grafico di ascissa $x_0 = 2$.
Dall'esempio 1.3 con $n = 3$ si ricava $f'(x) = 3x^2$.
Dalla formula (1.3), l'equazione della retta cercata è la seguente:

$$y = 8 + 12(x-2) = 12x - 16 .$$

La condizione di derivabilità è più stringente di quella di continuità, come mostra la seguente proposizione.

Proposizione 1.1 - *Sia $f:(a,b) \to \mathbb{R}$. Se f è derivabile in x allora è continua in x.*

Dimostrazione - Per $h \neq 0$, $x, x+h \in (a,b)$ si ha:

$$f(x+h) - f(x) = h \cdot \frac{f(x+h) - f(x)}{h} \longrightarrow 0 \cdot f'(x) = 0 \text{ per } h \to 0 .$$

Cioè $\lim_{h \to 0} f(x+h) = f(x)$, che prova la continuità di f in x. □

Il prossimo esempio, semplice ma *significativo*, indica che la proposizione 1.1 *non* si può invertire.

***Esempio* 1.5** - *Funzione continua ma non derivabile nell'origine.*
Sia $f(x) = |x|$. Essa è una funzione continua in $\mathbb{R}$.

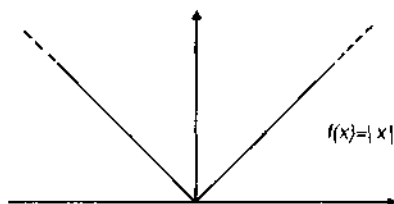



Fig. 6.2

Tuttavia si ha:

$$\frac{f(0+h)-f(0)}{h}=\frac{|h|-0}{h}=\frac{|h|}{h}\longrightarrow \pm 1 \text{ per } h\to 0_{\pm}\ .$$

Dunque il limite del rapporto incrementale per $h \to 0_+$ è diverso da quello per $h \to 0_-$. Si conclude che f *non* è derivabile in $x = 0$.

La figura 6.2 mostra che in $x = 0$ *non* è definita la retta tangente al grafico.

In un caso come quello dell'esempio 1.5 esistono *finiti* i limiti destro e sinistro del rapporto incrementale.

I due limiti

$$\lim_{h\to 0_+}\frac{f(x+h)-f(x)}{h} \quad \text{e} \quad \lim_{h\to 0_-}\frac{f(x+h)-f(x)}{h}$$

quando esistono finiti, si chiamano **derivata destra e sinistra** di f in x, rispettivamente, e si indicano con $f'_+(x)$ ed $f'_-(x)$.

Evidentemente f è derivabile in x se e solo se $f'_+(x) = f'_-(x)$.

Dal punto di vista geometrico si parlerà di tangente (o semitangente) dalla destra o dalla sinistra.

La proposizione 1.1 vale nella forma seguente: *se f è derivabile dalla destra in x (dalla sinistra) allora è continua dalla destra in x (dalla sinistra).*

Quando diciamo che f è derivabile in $[a, b]$, significa che f è derivabile in (a, b) e inoltre derivabile da destra in a, da sinistra in b.

Osservazione 1.1 - Dal significato geometrico di rapporto incrementale deduciamo che nei casi in cui f è continua e

$$\lim_{h\to 0}\frac{f(x_0+h)-f(x_0)}{h}=+\infty \text{ (oppure } -\infty)$$

esiste ancora la tangente al grafico di f nel punto $(x_0, f(x_0))$ ed è parallela all'asse delle ordinate (parleremo di *punti con tangente verticale*).

In questi casi, anche se f *non* è derivabile in x_0, scriveremo per comodità $f'(x_0) = +\infty$ (oppure $-\infty$).

Analogamente scriveremo $f'_+(x_0) = \pm\infty$ oppure $f'_-(x_0) = \pm\infty$ qualora il rapporto incrementale di f in x_0 abbia limite destro o sinistro infiniti.

a) Caso in cui $f'(x_0) = +\infty$ *b)* Caso in cui $f'(x_0) = -\infty$

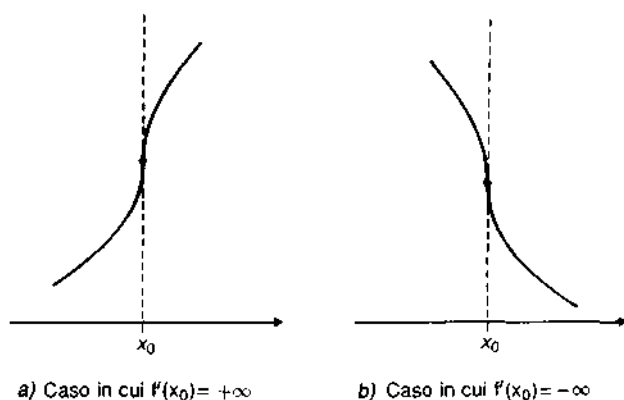



Fig. 6.3

Si chiamano *punti angolosi* i punti del grafico corrispondenti ad ascisse x in cui una funzione f è continua ma $f'_+(x) \neq f'_-(x)$. Una delle due derivate può anche essere $+\infty$ o $-\infty$.

Se $f'_+(x_0) = +\infty$ e $f'_-(x_0) = -\infty$, o viceversa, il punto si chiama *cuspide*.

a) P è un punto angoloso *b)* P è una cuspide: $f'_+ = -\infty$, $f'_- = +\infty$

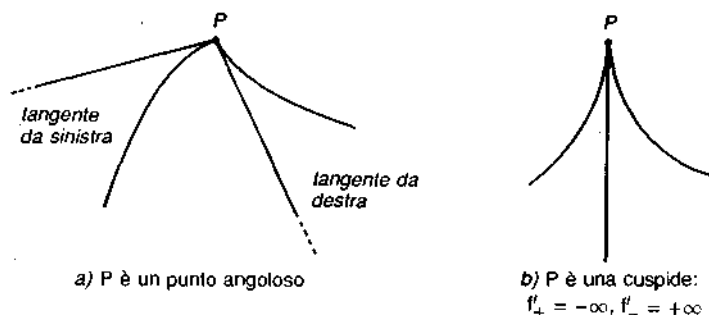



Fig. 6.4

Sia $f:(a,b) \to \mathbb{R}$, derivabile in (a,b). Se $f':(a,b) \to \mathbb{R}$ è a sua volta derivabile in (a,b), la sua derivata si indica con f'' oppure $D^2 f$, oppure $d^2 f/dx^2$ o $\ddot{f}$ e si chiama **derivata seconda** di f.

Se il processo può continuare n volte, è definita la **derivata** n**-esima** di f, indicata con uno dei simboli $f^{(n)}$, $D^n f$, $d^n f/dx^n$.

Si dice anche che $f^{(n)}$ è *la derivata di ordine* n di f. Per convenzione si pone $f^{(0)} = f$; ossia, la derivata di "ordine zero" di f è f stessa.

Concludiamo questo paragrafo con il calcolo delle derivate di alcune funzioni elementari.

Esempi

1.6. Sia $f(x) = x^\alpha(\alpha \in \mathbb{R})$. Per ogni $x \in \mathbb{R}_+$ abbiamo:

$$\frac{f(x+h)-f(x)}{h} = \frac{(x+h)^\alpha - x^\alpha}{h} = x^\alpha \cdot \frac{\left[\left(1+\frac{h}{x}\right)^\alpha - 1\right]}{h} .$$

Poiché, in base al limite (3.17) di 4.3.3 si ha:

$$\frac{\left[\left(1+\frac{h}{x}\right)^\alpha - 1\right]}{h} \to \frac{\alpha}{x} \text{ per } h \to 0$$

si conclude che $f'(x) = \alpha x^{\alpha-1}$.

Vale la pena di sottolineare alcuni casi particolari.

Se $\alpha = 1/2$, $f(x) = \sqrt{x}$, allora $f'(x) = \dfrac{1}{2\sqrt{x}}$; se $\alpha = 1/3$, $f(x) = \sqrt[3]{x}$, allora $f'(x) = \dfrac{1}{3\sqrt[3]{x^2}}$. Se $\alpha = -1$, $f(x) = 1/x$, allora $f'(x) = -1/x^2$.

1.7. Sia $f(x) = a^x \quad (a > 0, a \neq 1)$. Abbiamo, $\forall x \in \mathbb{R}$:

$$\frac{f(x+h)-f(x)}{h} = \frac{a^{x+h} - a^x}{h} = a^x \frac{a^h - 1}{h} \longrightarrow a^x \log a \text{ per } \; h \to 0$$

in base al limite (3.16) di 4.3.3.

Dunque $f'(x) = a^x \log a$.

1.8. Sia $f(x) = \log_a x \quad (a > 0, a \neq 1)$. Abbiamo, $\forall x \in \mathbb{R}_+$:

$$\frac{f(x+h)-f(x)}{h} = \frac{\log_a(x+h) - \log_a x}{h} = \frac{\log_a\left(1+\frac{h}{x}\right)}{h} \to \frac{1}{x \log a}$$

per $h \to 0$, in base al limite (3.15) di 4.3.3.

Dunque

$$f'(x) = \frac{1}{x \log a} .$$

1.9. Sia $f(x) = \sin x$; per ogni $x \in \mathbb{R}$ si ha, usando le formule di prostaferesi:

$$\frac{f(x+h)-f(x)}{h} = \frac{\sin(x+h) - \sin x}{h} = 2\frac{\sin\left(\frac{h}{2}\right)\cos\left(x+\frac{h}{2}\right)}{h} .$$

Essendo $\cos x$ continua, abbiamo

$$\cos\left(x+\frac{h}{2}\right) \longrightarrow \cos x \quad \text{per } h \to 0$$

e inoltre

$$\lim_{h\to 0} 2\frac{\sin\left(\frac{h}{2}\right)}{h} = \lim_{h\to 0} \frac{\sin\left(\frac{h}{2}\right)}{\frac{h}{2}} = 1 \; .$$

Dunque $f'(x) = \cos x$.

Analogamente si dimostra che la derivata di $\cos x$ è $-\sin x$.

Esercizi

1. Calcolare l'equazione della retta tangente al grafico di $f(x) = \sin x$ nel punto di ascissa $x = \pi/3$.

2. Verificare che $f(x) = \sqrt[3]{x^2}$ ha una cuspide nell'origine.

3. Determinare per quali valori di $a \in \mathbb{R}$, la seguente funzione è derivabile in $x = 0$:

$$f(x) = \begin{cases} x^a \sin \dfrac{1}{x} & x > 0 \\ 0 & x \leq 0 \; . \end{cases}$$

4. Sia $f: (a,b) \to \mathbb{R}$, derivabile in $x \in (a,b)$.
Calcolare il seguente limite in termini di $f'(x)$:

$$\lim_{h\to 0} \frac{f(x+h) - f(x-h)}{h} \; .$$

5. Sia $f(x) = a^x$; calcolare $f^{(n)}(x)$ per ogni $n \in \mathbb{N}$.

6. Calcolare $f^{(n)}(x)$ quando $f(x) = \sin x$ e quando $f(x) = \cos x$, per ogni $n \in \mathbb{N}$.

7. Sia f derivabile in x e $f'(x) = 0$. Dimostrare che

$$f(x+h) - f(x) = o(h) \text{ per } h \to 0 \; .$$

8. Determinare per quale valore di m la retta $y = mx$ è tangente al grafico di $y = e^x$.

9. Una funzione $f: [0,1] \to \mathbb{R}$, derivabile in ogni punto di $[0,1]$ ha il seguente grafico

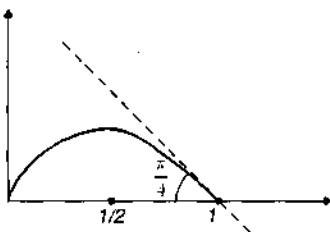



Fig. 6.5

Quale può essere il grafico della sua derivata tra i seguenti?

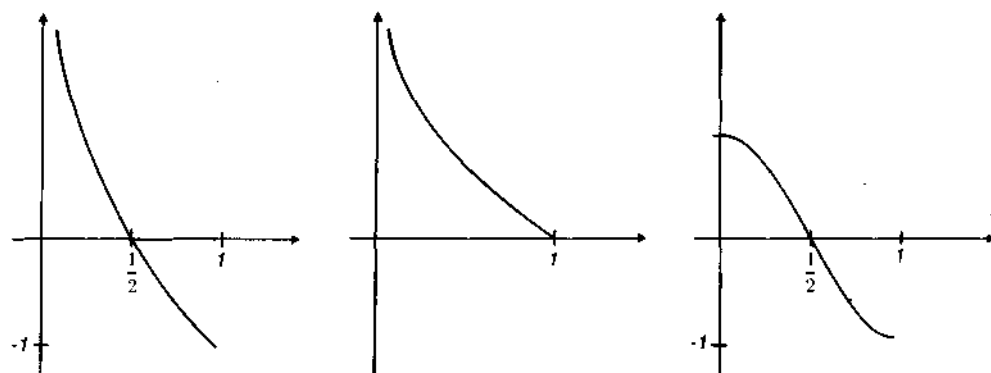



Fig. 6.6

10. Sia $f:(a,b) \to \mathbb{R}$ derivabile in $x_0 \in (a,b)$ e $f(x_0) = 0$. Dimostrare che $g = |f|$ è derivabile in x_0 se e solo se $f'(x_0) = 0$.

1.2 Algebra delle derivate

La derivazione può essere considerata come un'operazione che "produce" funzioni da altre funzioni.

In questo paragrafo esaminiamo la relazione tra derivazione e operazioni algebriche; esamineremo quindi (paragrafo 1.3) la relazione tra derivazione e composizione di funzioni, e poi (paragrafo 1.4) tra derivazione ed inversione di funzioni.

▪ **Teorema 1.2** - *Siano $f, g:(a,b) \to \mathbb{R}$, derivabili in $x \in (a,b)$. Allora $f+g$, $f \cdot g$ e f/g ($g(x) \neq 0$) sono derivabili in x e valgono le formule seguenti:*

i) $(f+g)'(x) = f'(x) + g'(x)$

ii) $(fg)'(x) = f'(x)g(x) + f(x)g'(x)$

iii) $(kf)'(x) = k \cdot f'(x)$ se k è costante

iv) $(f/g)'(x) = [f'(x)g(x) - f(x)g'(x)]/[g(x)]^2$.

Dimostrazione -

i) Sia $w = f + g$; allora

$$w(x+h) - w(x) = [f(x+h) - f(x)] + [g(x+h) - g(x)] .$$

Dividendo per h entrambi i membri e passando al limite per $h \to 0$ si ottiene la tesi.

ii) Sia $w = fg$; allora

$$\begin{aligned} w(x+h) - w(x) &= f(x+h)g(x+h) - f(x)g(x) = \\ &= g(x+h)[f(x+h) - f(x)] + f(x)[g(x+h) - g(x)] . \end{aligned}$$

Dividendo per h si ha

$$\frac{w(x+h)-w(x)}{h}=g(x+h)\frac{[f(x+h)-f(x)]}{h}+f(x)\frac{[g(x+h)-g(x)]}{h}\ . \quad (1.4)$$

Essendo g derivabile in x è anche continua, per cui $g(x+h)\to g(x)$ se $h\to 0$. Passando al limite per $h\to 0$ nella (1.4) si ottiene la tesi.

iii) Si deduce da *ii)* ponendo $g(x)=k$.

iv) Sia $w=f/g$; allora

$$\begin{aligned}w(x+h)-w(x)&=\frac{f(x+h)}{g(x+h)}-\frac{f(x)}{g(x)}=\\&=\frac{1}{g(x)\ g(x+h)}[f(x+h)\ g(x)-f(x)\ g(x+h)]=\\&=\frac{1}{g(x+h)\ g(x)}\ [g(x)\ (f(x+h)-f(x))-f(x)\ (g(x+h)-g(x))]\ .\end{aligned}$$

Dunque

$$\begin{aligned}&\frac{w(x+h)-w(x)}{h}=\\&=\frac{1}{g(x+h)\ g(x)}\left[g(x)\ \frac{f(x+h)-f(x)}{h}-f(x)\ \frac{g(x+h)-g(x)}{h}\right]\ .\end{aligned}$$

Passando al limite per $h\to 0$, notando ancora che $g(x+h)\to g(x)$ si ottiene la tesi. □

Esempi

1.10. Siano $f(x)=\cos x$, $g(x)=x^3$. Allora

$$\begin{aligned}&\frac{d}{dx}(\cos x+x^3)=\frac{d}{dx}\cos x+\frac{d}{dx}x^3=-\sin x+3x^2\\&\frac{d}{dx}(x^3\cos x)=\cos x\frac{d}{dx}x^3+x^3\frac{d}{dx}\cos x=\cos x\cdot 3x^2-x^3\sin x\\&\frac{d}{dx}\frac{\cos x}{x^3}=\frac{-\sin x\cdot x^3-3x^2\cos x}{x^6}=\frac{-\sin x\cdot x-3\cos x}{x^4}\end{aligned}$$

1.11. Sia $f(x)=\operatorname{tg} x=\sin x/\cos x$. Dalla *iv)* si ha:

$$\frac{d}{dx}\operatorname{tg} x\quad=\frac{(\sin x)^2+(\cos x)^2}{(\cos x)^2}=\frac{1}{(\cos x)^2}=1+(\operatorname{tg} x)^2\ .$$

1.12. Calcoliamo le derivate di $f(x) = x^n$.

$$f'(x) = nx^{n-1} \ , \ f''(x) = n(n-1)x^{n-2} \ ,$$
$$f'''(x) = n(n-1)(n-2)x^{n-3} \dots$$
$$\dots f^{(n)}(x) = n! \ , \ f^{(n+1)}(x) = f^{(n+2)}(x) = \dots = 0 \ .$$

Osservazione 1.2 - La $i)$ del teorema 1.2 si generalizza ad un numero n di addendi: se $f_1, \dots, f_n$ sono derivabili in x, allora

$$\left(\sum_{i=1}^{n} f_i\right)'(x) = \sum_{i=1}^{n} f_i'(x) \ .$$

Così pure la $ii)$ per un numero n di fattori:

$$(f_1 \cdot f_2 \cdot \ldots \cdot f_n)'(x) = \sum_{i=1}^{n} (f_1 \cdot f_2 \cdot \ldots \cdot f_i' \cdot \ldots \cdot f_n)(x) \ .$$

Consideriamo ora il prodotto di due funzioni f e g derivabili n volte in x. Dalle $i)$, $ii)$ del teorema 1.2 abbiamo:

$$\begin{aligned}
(fg)' &= f'g + fg' \\
(fg)'' &= f''g + f'g' + f'g' + fg'' = f''g + 2f'g' + fg'' \\
(fg)''' &= f'''g + f''g' + 2f''g' + 2f'g'' + f'g'' + fg''' = \\
&= f'''g + 3f''g' + 3f'g'' + fg''' \ .
\end{aligned}$$

Generalizzando si ottiene la formula di Leibniz seguente:

$$(fg)^{(n)} = \sum_{k=0}^{n} \binom{n}{k} f^{(n-k)} \, g^{(k)} \ . \tag{1.5}$$

Dalle $i)$ e $iii)$ del Teorema 1.2 ricaviamo che: la somma di funzioni derivabili e il prodotto di una funzione derivabile per un numero sono ancora funzioni derivabili, cosicché possiamo parlare di spazio vettoriale lineare delle funzioni derivabili.

Si introducono, a questo proposito, alcune notazioni particolarmente utili e comode.

Sia $I \subseteq \mathbb{R}$ un intervallo e $k \in \mathbb{N}$. Si indica con $C^k(I)$ lo spazio (vettoriale lineare) delle funzioni derivabili k volte in I e tali che la derivata k-esima sia continua; più brevemente diremo: lo *spazio delle funzioni k volte derivabili con continuità.*

Se $k = 0$ ritroviamo lo spazio $C^0(I)$ (indicato anche semplicemente con $C(I)$) delle funzioni continue in I.

Infine, $C^\infty(I)$ indica lo spazio delle funzioni che sono derivabili con continuità un numero qualsiasi di volte. Si ha evidentemente:

$$C^0(I) \supset C^1(I) \supset C^2(I) \supset \ldots \supset C^\infty(I) .$$

Se $f \in C^1(I)$ allora la sua derivata $f' \in C^0(I)$. L'operazione di derivazione definisce perciò una *corrispondenza univoca* dall'insieme $C^1(I)$ all'insieme $C^0(I)$, che indicheremo col simbolo D:

$$D: C^1(I) \to C^0(I) \ , \ f \mapsto Df = f'$$

(in accordo con la notazione già introdotta). Questa corrispondenza è perciò essa stessa una funzione (nel senso generale definito in 1.4.1); in questi casi (di corrispondenze tra insiemi di funzioni) è però meglio usare il termine *applicazione* o *operatore*: $C^1(I)$ è il dominio di definizione, $\mathrm{dom}(D)$, $C^0(I)$ è il codominio.

Possiamo anche assegnare all'operatore D un altro dominio di definizione, per esempio $C^2(I)$ oppure $C^\infty(I)$, ma non $C^0(I)$ (poiché non tutte le funzioni continue sono derivabili).

Le proprietà *i*) e *iii*) del teorema 1.2 si esprimono nella forma seguente:

$$\begin{aligned} &\forall f, g \in \mathrm{dom}(D) \\ &D(f+g) = Df + Dg \\ &D(kf) = kDf \quad (k \in \mathbb{R}) . \end{aligned}$$

Si vede così che D è un'*applicazione lineare* (cfr. 5.2.1) essendo additiva e omogenea.

1.3 Derivata di funzione composta. Derivata logaritmica

Sia $f: (a,b) \to \mathbb{R}$ derivabile in x; allora l'espressione

$$\varepsilon(h) := \frac{f(x+h) - f(x)}{h} - f'(x) \tag{1.6}$$

è infinitesima con h.

Dalla (1.6) ricaviamo, per $h \neq 0$

$$f(x+h) - f(x) = f'(x)h + h\varepsilon(h) . \tag{1.7}$$

Se definiamo $\varepsilon(0) = 0$, essendo $\varepsilon(h) \to 0$ per $h \to 0$, risulta che $\varepsilon(h)$ è continua in $h = 0$ e quindi la (1.7) continua a valere anche per $h = 0$.

Ritorneremo nella sezione 3 sul significato della (1.7); per ora ci servirà nella dimostrazione del prossimo teorema.

▪ **Teorema 1.3** - (Derivazione delle funzioni composte). *Siano* $f\colon(a,b)\to\mathbb{R}$, $g\colon(c,d)\to\mathbb{R}$ *con* $\operatorname{im}(f)\subseteq(c,d)$. *Se* f *è derivabile in* x *e* g *è derivabile in* $f(x)$, *la funzione composta* $w=g\circ f$ *è derivabile in* x *e vale la formula*

$$w'(x)=g'(f(x))\cdot f'(x)\,. \tag{1.8}$$

Prima di dare una dimostrazione del teorema premettiamo un'importante osservazione.

La formula (1.8) acquista maggior significato usando la notazione di Leibniz per le derivate.

Se poniamo

$$y=f(x)\quad \text{e}\quad w=g(y)$$

possiamo scrivere la (1.8) nella forma seguente:

$$\frac{dw}{dx}=\frac{dw}{dy}\cdot\frac{dy}{dx}\quad (\textit{come se } dy \text{ si semplificasse!}) \tag{1.8$'$}$$

La (1.8)' esprime il fatto che il tasso di variazione di w rispetto ad x è il prodotto dei tassi di variazione "intermedi", di w rispetto ad y e di y rispetto ad x.

Dimostrazione del teorema 1.3 - Dalla derivabilità di f in x e di g in $y=f(x)$ si ha, per ogni k sufficientemente piccolo (anche $k=0$, come abbiamo osservato prima di enunciare il teorema)

$$g(y+k)-g(y)=g'(y)k+k\varepsilon(k)\ \text{con}\ \varepsilon(k)\to 0\ \text{se}\ k\to 0\,. \tag{1.9}$$

Se poniamo $k=f(x+h)-f(x)$, abbiamo che, quando $h\to 0$,

$$k\to 0\quad \text{e}\quad \frac{k}{h}\to f'(x)\,.$$

Di conseguenza se $h\to 0$ anche $\varepsilon(k)\to 0$.

Dividendo la (1.9) per h, essendo $y+k=f(x+h)$

$$\frac{g(f(x+h))-g(f(x))}{h}=\frac{k}{h}[g'(f(x))+\varepsilon(k)]\,.$$

Passando al limite per $h\to 0$ si ottiene la (1.8). □

Esempi

1.13. Siano $f(x)=\sin x$ e $g(y)=y^3$.
Allora

$$w(x)=g(f(x))=(\sin x)^3\,.$$

Dalla (1.8)

$$w'(x) = 3(\sin x)^2 \cos x \ .$$

Usando la (1.8)$'$, si scriverebbe $y = \sin x$, $w = y^3$ e quindi

$$\frac{dw}{dx} = \frac{dw}{dy}\frac{dy}{dx} = 3y^2 \cos x = 3(\sin x)^2 \cos x \ .$$

1.14. Siano $f(x) = x^4 + 3x^2 + 1$ e $g(y) = \log y$.
Allora

$$w(x) = \log(x^4 + 3x^2 + 1) \ .$$

Dalla (1.8)

$$w'(x) = \frac{1}{x^4 + 3x^2 + 1}(4x^3 + 6x) \ .$$

Usando la (1.8)$'$ si pone $y = x^4 + 3x^2 + 1$ e $w = \log y$ e quindi:

$$\frac{dw}{dx} = \frac{dw}{dy}\frac{dy}{dx} = \frac{1}{y}\ (4x^3 + 6x) = \frac{1}{x^4 + 3x^2 + 1}(4x^3 + 6x) \ .$$

1.15. Un contenitore cilindrico con raggio di base $R = 1$ metri ed altezza 3 m è pieno d'acqua. Da un rubinetto posto in prossimità del fondo vengono prelevati 10 litri al minuto.

Con quale velocità l'altezza dell'acqua decresce?

Sia h l'altezza (in decimetri) della colonna d'acqua e V il suo volume (in dm^3).

Vogliamo trovare dh/dt, sapendo che $dV/dt = -10\,\mathrm{dm}^3\,/\,\mathrm{min}$.

Dalla (1.8)$'$:

$$\frac{dV}{dt} = \frac{dV}{dh} \cdot \frac{dh}{dt} \ ;$$

essendo $V = 100\pi h\,\mathrm{dm}^3$ otteniamo $dV/dh = 100\pi\,\mathrm{dm}^2$ ed infine

$$\frac{dh}{dt} = -\frac{10}{100\pi}\ \mathrm{dm}\,/\,\mathrm{min} = -\frac{1}{10\pi}\ \mathrm{dm}\,/\,\mathrm{min} \ .$$

Questo esempio mette in luce le possibilità di calcolo connesse con la formula (1.8)$'$.

Tramite il teorema 1.3, possiamo introdurre il concetto di derivata logaritmica di una funzione positiva.

Sia $f\colon (a, b) \to \mathbb{R}_+$, derivabile. Allora, in base al teorema 1.3 è derivabile anche la funzione $g = \log f$ e la sua derivata si chiama *derivata logaritmica di* f ed è data dalla formula

$$g'(x) = \frac{f'(x)}{f(x)} \,.$$

La derivata logaritmica, essendo il rapporto tra f' ed f, rappresenta il *tasso di variazione relativo di* f rispetto ad x.

Nei casi concreti tale tasso è spesso più significativo di f'.

Ad esempio, un incremento di un litro al minuto ha un effetto molto diverso se applicato a quantità di 10 o di 1000 litri. Nel primo caso il tasso di variazione relativo è 1/10, nel secondo 1/1000.

Per i motivi sopraesposti, quando interessa visualizzare gli incrementi relativi si ricorre a grafici in scala *semilogaritmica*: sull'asse delle ascisse si collocano i valori di x, mentre su quello delle ordinate quelli di $\log f(x)$. A questo tipo di rappresentazione si ricorre anche quando $f(x)$ cresce così rapidamente da richiedere una compressione troppo elevata della scala (unità di misura) sull'asse delle ordinate o in quella delle ascisse. Basti pensare al grafico di $f(x) = e^x$ che in scala semilogaritmica coincide con ... la bisettrice del 1° quadrante:

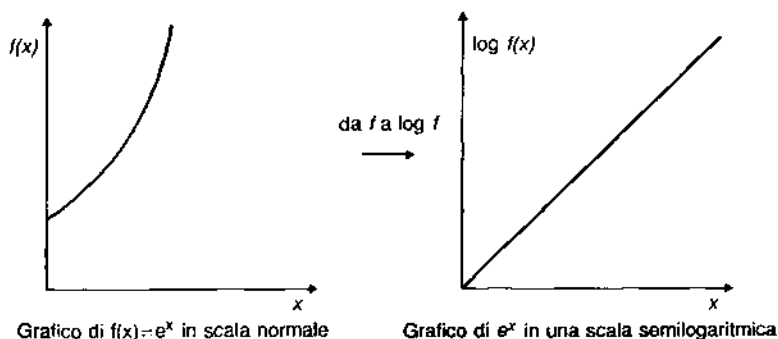



Fig. 6.7

Altra variante possibile è quella dei grafici in scala logaritmica, in cui invece di x sull'asse delle ascisse si collocano i valori di $\log x$; lo stesso si fa con f sull'asse delle ordinate.

La pendenza della retta tangente ad un grafico in scala logaritmica (cioè la derivata di $\log f$ rispetto a $\log x$) rappresenta il tasso di variazione relativa di f rispetto a variazioni relative di x; questa quantità prende il nome di *elasticità* di f e si indica con $E(x)$.

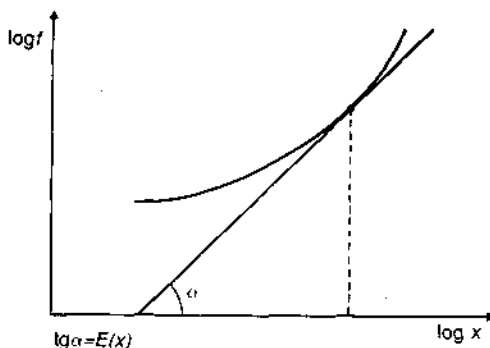



Fig. 6.8

Per trovare l'espressione analitica di $E(x)$, osserviamo che, per il teorema di derivazione delle funzioni composte, si ha, posto $u = \log x$:

$$\frac{d\log f}{dx} = \frac{d\log f}{du} \cdot \frac{du}{dx}$$

ovvero

$$\frac{f'(x)}{f(x)} = E(x)\frac{1}{x}$$

da cui si ricava

$$E(x) = x\frac{f'(x)}{f(x)} .$$

1.4 Derivata di funzione inversa

■ **Teorema 1.4** - (Derivazione delle funzioni inverse). *Sia $f:(a,b) \to \mathbb{R}$, continua e strettamente monotona. Se f è derivabile in $x_0 \in (a,b)$ e $f'(x_0) \neq 0$ allora la funzione inversa $g = f^{-1}$ è derivabile in $y_0 = f(x_0)$ e vale la formula*

$$g'(y_0) = \frac{1}{f'(x_0)} . \tag{1.10}$$

Con le notazioni di Leibniz, se poniamo $y = f(x)$ e $x = g(y)$ otteniamo dalla (1.10)

$$\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}} . \tag{1.10)'$$

Dimostrazione - Scriviamo il rapporto incrementale di g nella forma

$$\frac{g(y) - g(y_0)}{y - y_0} , \quad y \neq y_0 .$$

Ora, se $g(y) = x$ e $g(y_0) = x_0$, avremo, per definizione di funzione inversa, $y = f(x)$ e $y_0 = f(x_0)$ con $x \neq x_0$ poiché f è invertibile. Allora si può scrivere

$$\frac{g(y) - g(y_0)}{y - y_0} = \frac{x - x_0}{f(x) - f(x_0)} = \frac{1}{\frac{f(x) - f(x_0)}{x - x_0}} . \tag{1.11}$$

Essendo f continua in x_0 anche g lo è; dunque se $y \to y_0$, $x = g(y) \to g(y_0) = x_0$. Passando al limite per $y \to y_0$ nella (1.11), si ottiene la (1.10). □

***Esempio* 1.16** - Sia $f(x) = x + e^x$; f è continua e crescente (come somma di funzioni crescenti) in tutto $\mathbb{R}$, con immagine $\mathbb{R}$.

Sia $g = g(y)$ la funzione inversa. Calcoliamo $g'(1)$. Dalla formula (1.10) si ha, essendo $y_0 = 1$ l'immagine di $x_0 = 0$,

$$g'(1) = \frac{1}{f'(0)} = \frac{1}{2} .$$

Si noti che è impossibile trovare un'espressione esplicita di g; questa si dovrebbe ottenere risolvendo in x l'equazione $x + e^x = y$.

Esempi

1.17. Sia $f(x) = \sin x$, ristretta a $[-\pi/2, \pi/2]$. Calcoliamo la derivata di arcsin definita in 4.3.6.

Posto $y = \sin x$, abbiamo $f'(x) = \cos x = \sqrt{1 - (\sin x)^2} = \sqrt{1 - y^2}$; dalla (1.10):

$$g'(y) = \frac{1}{f'(x)} = \frac{1}{\sqrt{1 - y^2}} .$$

Dunque $\dfrac{d}{dy} \arcsin y = \dfrac{1}{\sqrt{1 - y^2}}$.

1.18. Sia $f(x) = \cos x$ ristretta a $[0, \pi]$. Calcoliamo la derivata di $g = \arccos$.

Posto $y = \cos x$, abbiamo $f'(x) = -\sin x = -\sqrt{1 - (\cos x)^2} = -\sqrt{1 - y^2}$. Dalla (1.10) si ha:

$$g'(y) = \frac{1}{f'(x)} = -\frac{1}{\sqrt{1 - y^2}} .$$

Dunque

$$\frac{d}{dy} \arccos y = \frac{-1}{\sqrt{1 - y^2}} .$$

1.19. Sia $f(x) = \operatorname{tg} x$ ristretta a $(-\pi/2, \pi/2)$ e $g = \operatorname{arctg}$.

Posto $y = \operatorname{tg} x$, si ha $f'(x) = 1 + (\operatorname{tg} x)^2 = 1 + y^2$; dalla (1.10) si ottiene

$$g'(y) = \frac{1}{f'(x)} = \frac{1}{1 + y^2} .$$

Dunque

$$\frac{d}{dy} \operatorname{arctg} y = \frac{1}{1 + y^2} .$$

1.20. Sia $f(x) = a^x$ $(a > 0, a \neq 1)$ e $g = \log_a$. Si ritrova subito

$$g'(y) = \frac{1}{y \log a} .$$

Concludiamo questa sezione con una tabella delle derivate di alcune funzioni elementari

f	f'	f	f'
$x^\alpha (\alpha \in \mathbb{R})$	$\alpha x^{\alpha-1}$	$\log_a x\ (a > 0, a \neq 1)$	$\dfrac{1}{x \log a}$
$\sin x$	$\cos x$	$\arcsin x$	$\dfrac{1}{\sqrt{1-x^2}}$
$\cos x$	$-\sin x$	$\arccos x$	$\dfrac{-1}{\sqrt{1-x^2}}$
$\operatorname{tg} x$	$1 + (\operatorname{tg} x)^2 = \dfrac{1}{(\cos x)^2}$	$\operatorname{arctg} x$	$\dfrac{1}{1+x^2}$
$a^x\ (a > 0)$	$a^x \log a$		

I teoremi contenuti nei paragrafi 1.2, 1.3, 1.4 indicano che la classe delle funzioni *elementari* è *chiusa rispetto all'operazione di derivazione.*

Esercizi

11. Sia $P(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$; calcolare $P^{(k)}(0)$ per $k = 0, \ldots, n$.

12. Qual è il tasso di variazione del volume di una sfera rispetto al suo raggio? e rispetto all'area della sua superficie?

13. Calcolare le derivate delle seguenti funzioni:

a) $\operatorname{cotg} x$ b) $\sec x$ c) $\operatorname{cosec} x$ d) $\operatorname{Sh} x$

e) $\operatorname{Ch} x$ f) $\operatorname{Th} x$ g) $\operatorname{SettSh} x$ h) $\operatorname{SettCh} x$ i) $\operatorname{SettTh} x$

l) $\dfrac{\sqrt{x}}{1+x^2}\ (x \geq 0)$ m) $(\operatorname{Sh} x)^3$

n) $e^{\sin x}$ o) $\operatorname{arctg}\left(\dfrac{x}{1-x^2}\right)\ (|x| \neq 1)$

p) $\log \operatorname{tg} x$ q) $\arcsin\left(\dfrac{1}{1+\sqrt{x}}\right)\ (x \geq 0)$

r) $x e^{-\frac{1}{x}}\ (x \neq 0)$ s) $2^{(\arccos 3x)}\ \left(|x| \leq \dfrac{1}{3}\right)$.

14. Determinare per quale valore del parametro reale a la seguente funzione ha derivata continua nell'origine

$$f(x) = \begin{cases} x^a \sin \dfrac{1}{x} & x > 0 \\ 0 & x \leq 0\,. \end{cases}$$

15. Usando l'identità $x = e^{\log x} (x > 0)$, calcolare la derivata di $f(x) = x^x$.

16. Generalizzando l'esercizio precedente trovare una formula per la derivata di

$$h(x) = [g(x)]^{f(x)}$$

dove g è una funzione positiva. Applicarla per calcolare la derivata di $f(x) = (1/x)^{x^2}$ $(x > 0)$.

17. In un triangolo isoscele come quello in figura, il vertice C si muove perpendicolarmente alla base in modo che l'area del triangolo cresca ad una velocità di 4 cm^2/sec.

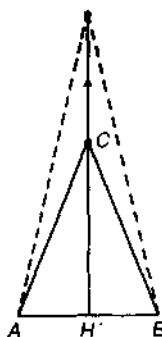



Fig. 6.9

La base del triangolo è lunga 3 cm. A quale velocità cresce l'altezza CH del triangolo? e a quale velocità cresce la distanza $\overline{CB}$?

18. Sia $f(x) = 2x + \log x (x > 0)$. Verificare che f è biiettiva tra $(0, +\infty)$ e $(-\infty, +\infty)$. Detta g la funzione inversa, calcolare $g'(2)$.

19. Nell'ipotesi del teorema 1.4, se f è due volte derivabile allora anche g lo è. Sfruttando la (1.10) nella forma

$$g'(y) = \frac{1}{f'(g(y))}$$

ed il teorema 1.3 trovare la formula per $g''(y)$.

20. Dimostrare che, se f è derivabile e $f(x) \neq 0$,

$$\frac{d}{dx} \log |f(x)| = \frac{f'(x)}{f(x)} .$$

21. Calcolare l'elasticità di x^a e a^x.

22. Sia $f: (a, b) \to \mathbb{R}$, derivabile in tutto (a, b).
Dimostrare che:

se f è pari, f' è dispari

se f è dispari, f' è pari

se f è periodica di periodo p, f' è periodica di periodo p.

1.5 Differenziale

Riprendiamo la formula (1.7). Se $f:(a,b) \to \mathbb{R}$ è derivabile in x, allora si può scrivere

$$f(x+h) - f(x) = f'(x)h + h\varepsilon(h) \tag{1.12}$$

dove $\varepsilon(h) \to 0$ con h.

Nella (1.12) l'incremento $\Delta f = f(x+h) - f(x)$ è scritto come somma di due termini:

$$f'(x)\,h \quad \text{lineare in } h$$

e

$$h\,\varepsilon(h) = o(h), \text{ infinitesimo di ordine superiore ad } h \text{ se } h \to 0\,.$$

Cambiando punto di vista, introduciamo la seguente

Definizione 1.2 - *Supponiamo che, dati $f:(a,b) \to \mathbb{R}$ ed un punto $x \in (a,b)$, esista un numero reale A tale che, per ogni h per cui $x+h \in (a,b)$, si abbia*

$$\Delta f = A\,h + o(h) \quad per \quad h \to 0\,. \tag{1.13}$$

Diciamo allora che f è differenziabile in x; la parte lineare Ah dell'incremento Δf si chiama **differenziale** *di f in x e si indica con il simbolo $df(x)$.*

Se nella (1.13) dividiamo per h entrambi i membri e passiamo al limite per $h \to 0$ otteniamo subito che f è derivabile in x e che $f'(x) = A$. Dunque

$$df(x) = f'(x)h\,.$$

Riassumendo, siano dati $f:(a,b) \to \mathbb{R}$ ed $x \in (a,b)$:

i) f derivabile in x significa che $\lim_{h\to 0} \Delta f/h$ esiste finito

ii) f differenziabile in x significa che Δf può essere approssimato da una parte lineare in h a meno di infinitesimi di ordine superiore rispetto ad h.

Dalla discussione precedente segue che *f è derivabile in x se e solo se è differenziabile in x*; inoltre il differenziale di f in x è dato dalla formula

$$df(x) = f'(x)h\,. \tag{1.14}$$

Vedremo nel cap. 7 che tale equivalenza è falsa per funzioni di più variabili.

Se calcoliamo il differenziale di $f(x) = x$ otteniamo

$$df(x) = dx = f'(x)h = h\,.$$

L'espressione (1.14) per il differenziale diventa dunque

$$df(x) = f'(x)dx \text{ (*)}$$

(*) Leibniz concepiva f' come rapporto delle quantità infinitesime df e dx!

L'algebra delle derivate si estende subito ai differenziali; se f e g sono differenziabili, valgono le formule seguenti:

$$d(f+g) = df + dg$$
$$d(fg) = gdf + fdg$$
$$d(f/g) = \frac{1}{g^2}(gdf - fdg) .$$

Riguardo alla composizione, il differenziale soddisfa una proprietà che si chiama *invarianza di forma* che lo rende più *flessibile* della derivata in alcune situazioni.

Siano $w = g(y)$ e $y = f(x)$ con g definita sull'immagine di f. La funzione w può allora essere riguardata come funzione della variabile y oppure come funzione della variabile x, attraverso la composizione $g \circ f$.

Nel 1° caso si ha:

$$dw = g'(y)dy ; \tag{1.15}$$

nel 2° caso si ha, utilizzando la regola di derivazione delle funzioni composte,

$$dw = g'(f(x)) \cdot f'(x)dx ,$$

che si può ancora scrivere

$$dw = g'(y)dy$$

essendo $y = f(x)$ e $dy = f'(x)dx$.

È sostanzialmente per l'invarianza di forma, che, nelle scienze applicate, si "differenzia" una data equazione anziché derivarla, quando non siano precisate le dipendenze delle variabili.

Ad esempio, in termodinamica, si può "differenziare" l'equazione

$$pV = RT \quad \text{(equazione di stato dei gas perfetti)}$$

ottenendo

$$pdV + Vdp = RdT ,$$

equazione valida qualunque sia la dipendenza da altre variabili di p, V, T.

Anche il differenziale può essere rappresentato geometricamente come viene mostrato in figura 6.10.

Come si vede, df è l'incremento in altezza valutato lungo la tangente invece che lungo la curva.

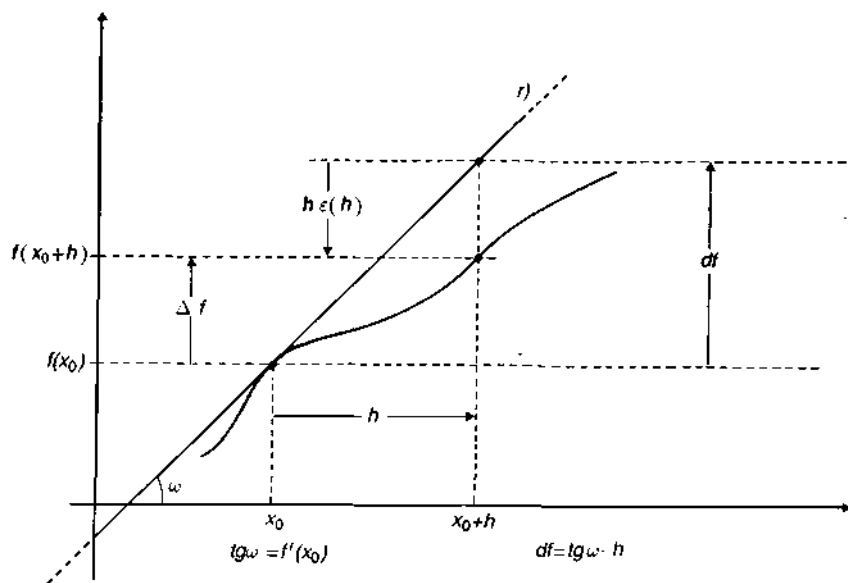



Fig. 6.10

Come si è fatto per le derivate, si possono introdurre anche i differenziali di ordine superiore; precisiamo subito che l'invarianza di forma vale *solo* per il differenziale primo, come il lettore potrà facilmente verificare.

Per una funzione derivabile n-volte in x, definiamo i differenziali secondo, terzo, $\ldots n$-esimo come segue:

$$d^2f(x) := f''(x)dx^2 \ , \ d^3f(x) := f'''(x)dx^3 \ , \ \ldots \ ,$$
$$d^nf(x) := f^{(n)}(x)dx^n \ .$$

Esercizi

23. Calcolare l'espressione del differenziale secondo di $f(x)$ quando si pensi x variabile indipendente oppure quando $x = x(t)$. Verificare che non vale l'invarianza di forma.

24. Estendere la formula di Leibniz (1.5) ai differenziali.

2. I TEOREMI FONDAMENTALI DEL CALCOLO DIFFERENZIALE

2.1 Teorema di Fermat. Estremi locali

I teoremi presentati in questa sezione forniscono gli strumenti principali ed indispensabili per lo studio delle funzioni da $\mathbb{R}$ in $\mathbb{R}$.

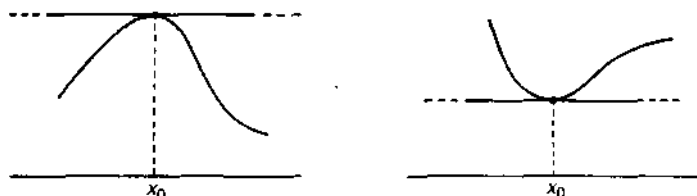



Fig. 6.11

Il primo di questi, sostanzialmente dovuto al matematico francese Pierre de Fermat (fine del 17° secolo), esprime un fatto geometricamente intuitivo: se esiste la tangente in un punto $(x_0, f(x_0))$ del grafico di una funzione e $f(x_0)$ è un estremo locale (massimo o minimo) allora essa è parallela all'asse delle ascisse.

▪ **Teorema 2.1** - (di Fermat). *Sia $f:(a,b) \to \mathbb{R}$ ed $x_0 \in (a,b)$. Se in x_0 f è derivabile ed f ha un estremo locale in x_0 allora $f'(x_0) = 0$.*

Dimostrazione - Sia x_0 punto di massimo locale. Per ipotesi esiste un intorno I di x_0, $I \subset (a,b)$, tale che $f(x_0) \geq f(x)$, $\forall x \in I$.

Consideriamo il rapporto incrementale

$$\Phi(x) = \frac{f(x) - f(x_0)}{x - x_0} \quad , \quad x \neq x_0 \ .$$

Per $x \in I$, si ha:

$$\begin{aligned} x > x_0 \ &\Rightarrow \ \Phi(x) \ \leq 0 \\ x < x_0 \ &\Rightarrow \ \Phi(x) \ \geq 0 \ . \end{aligned}$$

Per il teorema della permanenza del segno si deduce che:

$$\lim_{x \to x_0+} \Phi(x) = f'_+(x_0) \leq 0 \ \text{ e } \ \lim_{x \to x_0-} \Phi(x) = f'_-(x_0) \geq 0 \ .$$

Essendo f derivabile in x_0, si conclude che $f'_+(x_0) = f'_-(x_0) = 0$. Analogamente si procede quando x_0 è punto di minimo locale. □

Definizione 2.1 - *I punti x in cui una funzione f ha derivata nulla si dicono* **punti stazionari** *o* **critici** *(*).*

Il teorema 2.1 afferma che *se x_0 è punto di estremo per f, interno al dominio di f, ed f è ivi derivabile, allora x_0 è stazionario.*

Naturalmente non tutti i punti stazionari per una funzione sono di estremo.

Ad esempio $f(x) = x^3$ ha in 0 un punto stazionario che *non* è estremo.

Punti come l'origine per la funzione x^3 si chiamano punti di flesso o di inflessione; ne parleremo in dettaglio nel paragrafo 3.1.

(*) Osserviamo che alcuni autori definiscono punti critici le immagini dei punti stazionari.

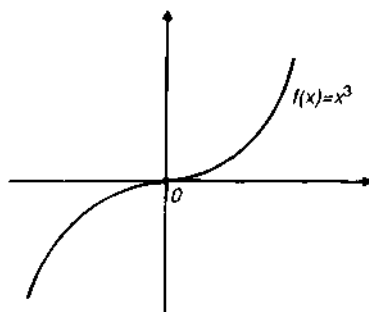



Fig. 6.12

Dal punto di vista pratico il teorema 2.1 costituisce il primo passo nella soluzione del problema dell'ottimizzazione di una funzione reale di una variabile reale, della ricerca cioè dei suoi estremi, locali o globali. Esso infatti, come abbiamo già osservato, indica che i punti di estremo in cui una funzione f è derivabile si trovano tra le soluzioni dell'equazione $f'(x) = 0$, cioè tra i punti stazionari di f.

Il secondo passo consisterà nell'esaminare la natura di un punto stazionario. A questo proposito risultano decisivi i prossimi teoremi e, in particolare, le conseguenze del teorema di Lagrange.

2.2 Teoremi di Rolle, Cauchy, Lagrange

Cominciamo con il teorema di Rolle (Michel Rolle, 1652-1719), il quale stabilisce condizioni *minimali* per l'esistenza di un punto stazionario per funzioni definite in intervalli *chiusi* e *limitati*.

■ **Teorema 2.2** - (di Rolle). *Sia* $f: [a, b] \to \mathbb{R}$, *soddisfacente le seguenti condizioni:*

i) f *è continua in* $[a, b]$,

ii) f *è derivabile in* (a, b),

iii) $f(a) = f(b)$.

Allora esiste $c \in (a, b)$ *tale che* $f'(c) = 0$.

Dimostrazione - Dal teorema di Weierstrass sappiamo che f ha massimo e minimo (globali); siano dunque $x_0, x_1 \in [a, b]$ tali che

$$f(x_0) = M = \max_{([a,b])} f$$

$$f(x_1) = m = \min_{([a,b])} f \, .$$

Se $M = m$ allora f è costante e $f'(x) = 0$ per ogni $x \in (a, b)$.

Se $M > m$, dalla *iii)* deduciamo che *almeno* uno tra i due punti x_0, x_1 è *interno*. In quel punto la derivata esiste per la *ii)* e deve essere nulla per il teorema di Fermat. □

$f(x) = \begin{cases} x & x \in [0,1) \\ 0 & x = 1 \end{cases}$
Valgono *ii*), *iii*) ma non *i*)

$f(x) = x, x \in [0,1]$
Valgono *i*), *ii*) ma non *iii*)

$f(x) = |x|, x \in [-1,1]$
Valgono *i*) e *iii*) ma non *ii*)

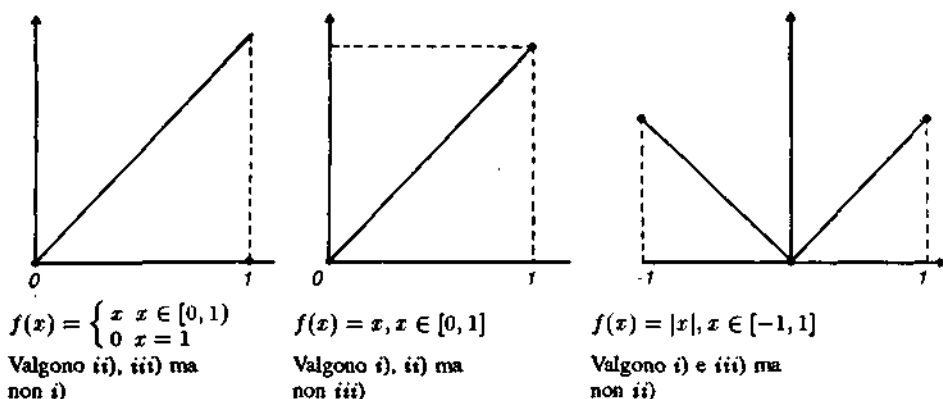



Fig. 6.13

Abbiamo detto che le ipotesi del teorema sono minimali. Il significato di questa frase è che, se indeboliamo una sola di esse, il teorema cessa di valere come mostrano gli esempi illustrati dalla figura 6.13.

Come si vede, in tutti e tre gli esempi non esiste alcun punto del grafico di f con tangente orizzontale.

Dal teorema di Rolle deduciamo i due importanti teoremi seguenti.

▪ **Teorema 2.3** - (di Cauchy). *Siano $f, g: [a,b] \to \mathbb{R}$, soddisfacenti le seguenti condizioni:*

i) *f, g continue in $[a,b]$,*

ii) *f, g derivabili in (a,b).*

Allora esiste $c \in (a,b)$ tale che

$$[f(b) - f(a)]\, g'(c) = [g(b) - g(a)]\, f'(c)\ . \tag{2.1}$$

Dimostrazione - Definiamo $w(x) = [f(b) - f(a)]\, g(x) - [g(b) - g(a)]\, f(x)$.
Allora w è continua in $[a,b]$, derivabile in (a,b) e $w(a) = w(b)$. Per il teorema di Rolle esiste $c \in (a,b)$ tale che $w'(c) = [f(b) - f(a)]\, g'(c) - [g(b) - g(a)] f'(c) = 0$. □

Osservazione 2.1 - Se $g'(x) \neq 0$ in (a,b), allora deve essere $g(b) \neq g(a)$ (per il teorema di Rolle) e la (2.1) si può scrivere nella forma

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(c)}{g'(c)}\ . \tag{2.2}$$

Scegliendo nella (2.2) $g(x) = x$ otteniamo il seguente importantissimo

▪ **Teorema 2.4** - (del valor medio, di Lagrange). *Sia $f: [a,b] \to \mathbb{R}$, continua in $[a,b]$ e derivabile in (a,b). Allora esiste un punto $c \in (a,b)$ ("punto di Lagrange") tale che:*

$$\frac{f(b) - f(a)}{b - a} = f'(c)\ . \tag{2.3}$$

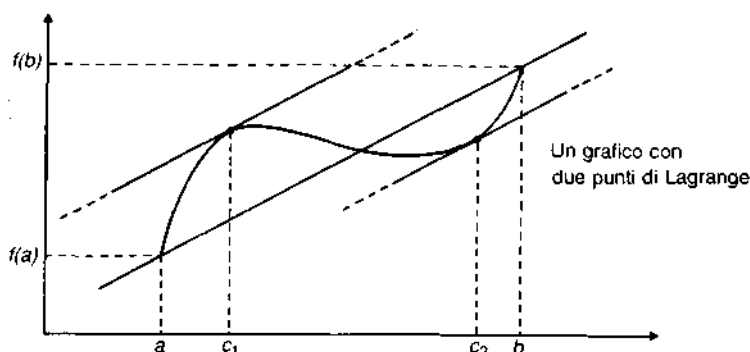



Fig. 6.14

Il significato geometrico della (2.3) è evidente: nelle condizioni del teorema esiste (almeno) un punto $(c, f(c))$ sul grafico in cui la tangente è parallela alla retta passante per i punti $(a, f(a))$ e $(b, f(b))$.

La (2.3) si può applicare ad un intervallo $[x, x+h] \subseteq [a, b]$; scrivendo allora $x+h$ al posto di b, x al posto di a, e $x+\theta h$ (con $0 < \theta < 1$) al posto di c, la (2.3) può mettersi nella forma:

$$f(x+h) = f(x) + hf'(x+\theta h) \qquad (0 < \theta < 1) \tag{2.3'}$$

Naturalmente θ viene a dipendere da x e da h.

***Esempio* 2.1** - Determiniamo i punti di Lagrange per la funzione $f(x) = x + \operatorname{arctg} x$ nell'intervallo $[0, 1]$.

Evidentemente f verifica le ipotesi del teorema 2.4, quindi i punti c di Lagrange sono soluzioni dell'equazione

$$f'(x) = \frac{f(1) - f(0)}{1 - 0}$$

e cioè

$$1 + \frac{1}{1+x^2} = \frac{\pi}{4} + 1 \, .$$

da cui l'unico punto

$$x = \sqrt{\frac{4}{\pi} - 1} \, .$$

2.3 Prime conseguenze del teorema di Lagrange

È chiaro che la determinazione dei punti c può non essere semplice; comunque la sola esistenza di tali punti determina le seguenti conseguenze:

a) *Test di monotonia per funzioni derivabili in un intervallo.*

Sia $f:(a,b) \to \mathbb{R}$, derivabile in (a,b). Allora f è crescente in (a,b) se e solo se $f'(x) \geq 0$, $\forall x \in (a,b)$.

Dimostrazione - Sia $f'(x) \geq 0$ in (a,b) e siano $x_2, x_1 \in (a,b)$ con $x_2 > x_1$; dobbiamo dimostrare che $f(x_2) \geq f(x_1)$. Applichiamo il teorema del valor medio ad f nell'intervallo $[x_1, x_2]$, dove le ipotesi sono chiaramente verificate. Si ottiene che esiste $c \in (x_1, x_2)$ tale che:

$$f(x_2) - f(x_1) = f'(c)(x_2 - x_1) .$$

Essendo $f'(c) \geq 0$, $x_2 - x_1 > 0$ si deduce $f(x_2) \geq f(x_1)$. Viceversa, se f è crescente in (a,b), allora, $\forall x, x+h \in (a,b)$ risulta $[f(x+h) - f(x)]/h \geq 0$, da cui segue $f'(x) \geq 0$ □

Si ha anche

$$f'(x) \leq 0 \quad \Leftrightarrow \quad f \text{ decrescente}$$
$$f'(x) < 0 \quad \Rightarrow \quad f \text{ strettamente decrescente}$$
$$f'(x) > 0 \quad \Rightarrow \quad f \text{ strettamente crescente}$$

con dimostrazione identica.

b) *Test di riconoscimento dei punti stazionari.*

Sia $f:(a,b) \to \mathbb{R}$, $x_0 \in (a,b)$ ed $f'(x_0) = 0$. Se esistono un intorno destro (sinistro) in cui $f' > 0$ ed un intorno sinistro (destro) in cui $f' < 0$, $(x \neq x_0)$, allora x_0 è un punto di minimo (massimo) locale forte.

Dimostrazione - Ovvia da a). Graficamente:

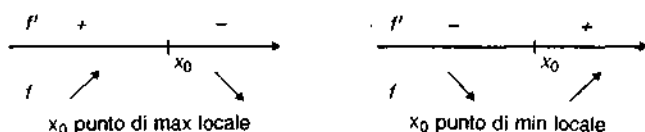



Fig. 6.15

***Esempio* 2.2** - Sia $f(x) = xe^{-\frac{1}{2}x^2}$; poiché $f'(x) = (1 - x^2)e^{-\frac{1}{2}x^2}$, gli unici punti stazionari sono $x = 1$ e $x = -1$.

Essendo $1 - x^2 > 0$, per $-1 < x < 1$, si ricava che $x = 1$ è di massimo forte mentre $x = -1$ è di minimo forte.

c) *Sia $f:(a,b) \to \mathbb{R}$, derivabile in (a,b). Allora f = costante se e solo se $f'(x) = 0$, $\forall x \in (a,b)$.*

Dimostrazione - Se f = costante allora $f'(x) = 0$ $\forall x \in (a,b)$. Viceversa se $f'(x) = 0$ $\forall x \in (a,b)$ il ragionamento del punto a) indica che $f(x_2) = f(x_1)$ per ogni coppia $x_2, x_1 \in (a,b)$, e cioè che f è costante. □

d) *Siano $f,g:(a,b) \to \mathbb{R}$ derivabili in (a,b). Se $f'(x) = g'(x)$ $\forall x \in (a,b)$ allora* $f = g +$ costante.

Dimostrazione - Porre $w = f - g$ e applicare c). □

La conseguenza d) dà informazioni sull'insieme delle *primitive* di una funzione in un *intervallo.*

Definizione 2.2 - *Sia* $\varphi: (a,b) \to \mathbb{R}$; *una funzione* f *si dice* **primitiva** *di* φ *in* (a,b) *se* f *è derivabile in* (a,b) *e* $f'(x) = \varphi(x)\ \forall x \in (a,b)$.

La d) implica che *se* φ *ha una primitiva* f *in* (a,b), allora *ne ha infinite* e *tutte* sono date dalla formula: $f(x)$+costante.

***Esempio* 2.3** - Sia $\varphi(x) = x^3$; poiché $f(x) = \dfrac{1}{4}x^4$ è derivabile e $f'(x) = x^3$ si ottiene che tutte le primitive di φ in $\mathbb{R}$ sono tutte e sole le funzioni: $\dfrac{1}{4}x^4 + k$ $(k \in \mathbb{R})$.

Ritorneremo ampiamente sul concetto di primitiva e sul calcolo delle eventuali primitive di una data funzione nel capitolo dedicato all'integrazione.

Con la frase ... *se* φ *ha una primitiva* ... abbiamo implicitamente posto la questione se ci siano funzioni che *non* ammettono primitiva, ovvero che *non* possono essere *derivate di un'altra funzione.*

Una risposta parziale è data dalla seguente proposizione:

e) *Proprietà dei valori intermedi per le derivate.*

Siano $f: (a,b) \to \mathbb{R}$, *derivabile in* (a,b) *e* $x_1, x_2 \in (a,b)$. *Allora* f' *assume tutti i valori compresi tra* $f'(x_1)$ *ed* $f'(x_2)$.

Dimostrazione - Sia ad esempio $f'(x_1) < f'(x_2)$ e $f'(x_1) < \lambda < f'(x_2)$.
Dobbiamo mostrare che esiste $c \in (x_1, x_2)$ tale che $f'(c) = \lambda$. Poniamo $g(x) = f(x) - \lambda x$. Osserviamo che $g'(x) = f'(x) - \lambda$ e quindi che $g'(x_1) = f'(x_1) - \lambda < 0$ e $g'(x_2) = f'(x_2) - \lambda > 0$. Deduciamo che g *non* è *monotona* in $[x_1, x_2]$, poiché la sua derivata cambia segno. Allora esistono due numeri $\alpha, \beta \in [x_1, x_2]$, $\alpha < \beta$, tali che $g(\alpha) = g(\beta)$; nell'intervallo $[\alpha, \beta]$ g soddisfa le ipotesi del teorema di Rolle per cui esiste $c \in (\alpha, \beta)$ tale che $g'(c) = 0$, che vuol dire $f'(c) - \lambda = 0$. □

Da e) si deduce ad esempio che una funzione con una o più *discontinuità* di 1^a *specie* in un intervallo, *non* può avere primitiva.

***Esempio* 2.4** - Sia:

$$\varphi(x) = \begin{cases} 1 & \text{per } x \geq 0 \\ 0 & \text{per } x < 0 \end{cases} \quad \text{("gradino" di Heaviside)} .$$

Se fosse $f'(x) = \varphi(x)$ in un intervallo contenente $x = 0$ avremmo che f' *non* assume i valori tra 0 ed 1, contro la e). Dunque φ non ha primitiva in ogni intervallo contenente l'origine.

Osservazione 2.2 - Se f è derivabile in (a, b), *f' può avere discontinuità di 2^a specie:*

Esempio 2.5 - Se

$$f(x) = \begin{cases} x^2 \sin \dfrac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases}$$

si ha:

$$f'(x) = \begin{cases} 2x \sin \dfrac{1}{x} - \cos \dfrac{1}{x} & x \neq 0 \\ 0 & x = 0 \ . \end{cases}$$

Si vede dunque che $\lim\limits_{x \to 0} f'(x)$ non esiste e perciò f' ha una discontinuità di 2^a specie in $x = 0$.

f) *Sulla funzione esponenziale*

Riprendiamo in esame la funzione esponenziale $x \mapsto a^x$ $(a > 0, a \neq 1)$. Abbiamo visto in 5.3.4 come essa sia "caratterizzata" da una *equazione funzionale*, precisamente dall'equazione $f(x + y) = f(x)f(y)$. Vediamo ora che tale funzione possiede un'altra proprietà caratteristica, espressa dalla cosiddetta *equazione di crescita* (o di decadimento):

$$f'(x) = kf(x) \quad (k \in \mathbb{R}) \ . \tag{2.4}$$

Osserviamo subito che, se $f(x) = a^x$, la (2.4) è soddisfatta con $k = \log a$.

La (2.4) è un'equazione nell'incognita f; questa incognita compare insieme alla sua derivata f'; per questo motivo la (2.4) viene detta *equazione differenziale*.

Tali tipi di equazioni, estremamente importanti, saranno trattate estesamente nel secondo volume.

Tornando alla (2.4), essa esprime che il "tasso di crescita di f è proporzionale ad f, nello stesso istante" (se interpretiamo x come tempo).

Questa legge è propria dell'evoluzione dinamica, ad esempio, del decadimento radioattivo in cui x è il tempo ed $f(x)$ rappresenta la quantità di materia radioattiva rimasta al tempo x.

In questo caso, trattandosi di decadimento, la costante k di proporzionalità della (2.4) è negativa e dipende dal tipo di sostanza in esame.

Vale il seguente teorema:

■ **Teorema 2.5** - *Se $f: \mathbb{R} \to \mathbb{R}$ è derivabile, $f(0) = 1$ e f soddisfa (2.4), allora* $f(x) = e^{kx}$.

Dimostrazione - Facciamo vedere che f soddisfa l'equazione funzionale $f(x + y) = f(x)f(y)$ che può essere riscritta nella forma:

$$f(y - x)f(x) = f(y) \quad x, y \in \mathbb{R} \ . \tag{2.5}$$

Consideriamo la funzione $\varphi(x) = f(y-x)f(x)$ con y fissato. Allora φ è derivabile e

$$\begin{aligned}\varphi'(x) &= -f'(y-x)f(x) + f(y-x)f'(x) = \text{ (usando la (2.4))}\\ &= -kf(y-x)f(x) + kf(y-x)f(x) = 0\end{aligned}$$

per ogni $x \in \mathbb{R}$. Dunque $\varphi(x)$ è costante ovvero $\varphi(x) = \varphi(0)$, $\forall x \in \mathbb{R}$, ovvero $f(y-x)f(x) = f(y)f(0) = f(y)$ essendo $f(0) = 1$. Ne segue che f soddisfa la (2.5) e quindi $f(x) = a^x = e^{x \log a}$.

Sostituendo l'ultima espressione nella (2.4) si ricava

$$e^{x \log a} \log a = k e^{x \log a}$$

da cui $k = \log a$. □

Esercizi

1. Sia $f\colon [a,b] \to \mathbb{R}$. Supponiamo che $x_0 \in [a,b]$ sia punto di estremo locale e che in x_0 f sia derivabile (si intende che se $x_0 = a$ o $x_0 = b$ la derivabilità è solo dalla destra o dalla sinistra, rispettivamente).

Dimostrare che se x_0 è di massimo allora $f'(x_0)(x - x_0) \leq 0$ mentre se x_0 è di minimo allora $f'(x_0)(x - x_0) \geq 0$.

2. Sia f continua in $[a,b)$, derivabile in (a,b). Dimostrare, usando il teorema di Lagrange, che se $\lim_{x \to a^+} f'(x) = r$, $r \in \mathbb{R}$, allora $r = f'_+(a)$.

3. Siano

$$f(x) = \frac{x}{x^2-1}\ ,\ f(x) = \frac{\log x}{x}\ ,\ f(x) = xe^{-\frac{1}{x}}\ .$$

Determinare gli eventuali punti stazionari e le zone di crescenza e decrescenza. Dedurre la natura dei punti stazionari. Determinare se le funzioni hanno estremi assoluti nel loro dominio.

4. Sia $f\colon (a,b) \to \mathbb{R}$; f si dice *lipschitziana* in (a,b) se il suo rapporto incrementale relativo ad una qualunque coppia di punti si mantiene limitato: cioè se $\forall x_1, x_2 \in (a,b)$ esiste M tale che

$$\frac{|f(x_1) - f(x_2)|}{|x_1 - x_2|} \leq M\ .$$

Un esempio di funzione lipschitziana in $\mathbb{R}$ è $f(x) = |x|$.

Dimostrare che se f è derivabile e $|f'(x)| \leq M$ $\forall x \in (a,b)$ allora f è lipschitziana in (a,b).

5. Dimostrare che $\forall x \in \mathbb{R}$, si ha $|\sin x - \sin y| \leq |x - y|$.

6. Sia f derivabile in $\mathbb{R}$. Dimostrare che f' porta connessi in connessi.

7. Un oggetto viene lanciato verso l'alto in verticale con una velocità v_0. Quale altezza massima raggiunge?

8.* (Riflessione della luce). Un raggio luminoso collega due punti A e B dopo essere stato riflesso da uno specchio (vedi figura). Determinare il punto di riflessione utilizzando il "principio di Fermat" dell'ottica geometrica (*il percorso "scelto" dal raggio riflesso sarà*

quello che richiede il tempo di percorrenza minore). Ritrovare la legge della riflessione: $i = r$.

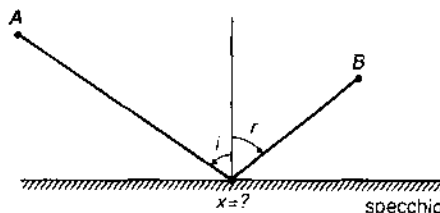



Fig. 6.16

9.* (Rifrazione della luce). Un raggio luminoso passa da un mezzo dove viaggia con velocità c_1 ad un altro mezzo nel quale viaggia con velocità c_2, separato dal primo da un piano, connettendo due punti A e B. Determinare il punto di rifrazione del raggio, utilizzando il principio di Fermat dell'esercizio precedente.

Ritrovare la legge di Snell per la rifrazione: $\dfrac{c_1}{c_2} = \dfrac{\sin i}{\sin r}$.

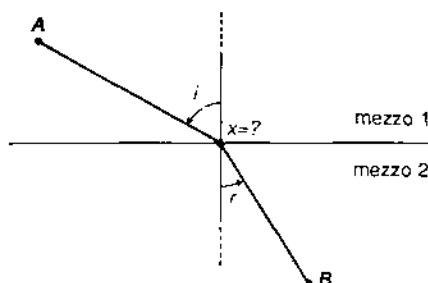



Fig. 6.17

10. Verificare che i punti di Lagrange per le funzioni $f(x) = x^2$ ed $f(x) = 1/x$ in un intervallo $[a, b]$ (con $0 < a < b$ per la seconda funzione) concidono rispettivamente con la media aritmetica $\dfrac{a+b}{2}$ e la media geometrica $\sqrt{ab}$ dei due numeri a e b.

11. Verificare se le seguenti successioni sono definitivamente monotone

$$a_n = n - \sqrt{n^2 + 1} \ , \ a_n = ne^{-n} \ , \ a_n = n \ \sin n \ .$$

12. Sia $f: \mathbb{R} \to \mathbb{R}$ derivabile con continuità e siano $f(2) = 1$, $f'(2) = 4$. Dimostrare che f è invertibile in un intorno di $x = 2$ e calcolare $(f^{-1})'(1)$.

13. Quale tra i due numeri π^e e e^π è il più grande? (senza usare calcolatrici...).

2.4 Il teorema di de L'Hôpital

L'argomento di questo paragrafo è un teorema usato frequentemente per calcolare limiti che presentano la forma di indecisione $0/0$ o ∞/∞.

▪ **Teorema 2.6** - (di de L'Hôpital) (*). *Siano* $-\infty \le a < b \le +\infty$ *e* $f, g: (a,b) \to \mathbb{R}$, *soddisfacenti le seguenti condizioni:*

i) $\lim_{x \to a_+} f(x) = \lim_{x \to a_+} g(x) = 0$ *oppure* $+\infty, -\infty, \infty$.

ii) f, g *derivabili in* (a,b) *e* $g'(x) \ne 0$, $\forall x \in (a,b)$.

iii) $\lim_{x \to a_+} \dfrac{f'(x)}{g'(x)} = L$ *(finito o infinito).*

Allora

$$\lim_{x \to a_+} \frac{f(x)}{g(x)} = L \; .$$

Le stesse conclusioni valgono se $x \to b_-$. Notiamo che, essendo $g'(x) \ne 0$, g è strettamente monotona e quindi diversa da zero in un intorno destro di a (o sinistro di b); dunque f/g è ben definita in questo intorno.

Osservazione 2.3 - La ragione del "legame" tra i limiti di f/g e di f'/g' si vede bene nel caso $0/0$ ed a finito.

Infatti, in tal caso si può definire $f(a) = g(a) = 0$; f e g risultano continue in $[a,b)$ e si può scrivere, per $x \ne a$:

$$\frac{f(x)}{g(x)} = \frac{f(x) - f(a)}{g(x) - g(a)} = \frac{f(x) - f(a)}{x - a} \frac{x - a}{g(x) - g(a)} \; .$$

Se supponiamo che $f'_+(a)$ ed $g'_+(a)$ esistano e $g'_+(a) \ne 0$, passando al limite per $x \to a_+$ si ha:

$$\lim_{x \to a_+} \frac{f(x)}{g(x)} = \frac{f'_+(a)}{g'_+(a)} \; .$$

Se ulteriormente f' e g' fossero continue dalla destra in a avremmo finalmente

$$\lim_{x \to a_+} \frac{f(x)}{g(x)} = \lim_{x \to a_+} \frac{f'(x)}{g'(x)} \; .$$

Osservazione 2.4 - Sottolineiamo che le sole ipotesi *i)* e *ii)* *non* garantiscono che $f(x)/g(x)$ e $f'(x)/g'(x)$ abbiano lo stesso comportamento quando $x \to a_+$. È essenziale la *iii)* come mostra il seguente semplice esempio.

Siano $f(x) = 2x + \sin x$, $g(x) = 2x - \sin x$.

(*) Guillaume de L'Hôpital (1661-1704), autore del primo trattato di calcolo infinitesimale.

Sia f sia g soddisfano le ipotesi $i)$ ed $ii)$ del teorema in un intorno di $+\infty$, ad esempio in $(2, +\infty)$.

Osserviamo che

$$\lim_{x \to +\infty} \frac{f'(x)}{g'(x)} = \lim_{x \to +\infty} \frac{2 + \cos x}{2 - \cos x}$$

non esiste. Tuttavia

$$\lim_{x \to +\infty} \frac{2x + \sin x}{2x - \sin x} = 1 \ .$$

La dimostrazione del teorema è rimandata alla fine del paragrafo; vediamo prima alcuni esempi di applicazione.

Esempi

2.6. Sia da calcolare $\lim_{x \to 0} (\sin x - x)/x^3$ che presenta una forma di indecisione $0/0$.

Si ha:

$$\frac{f'(x)}{g'(x)} = \frac{\cos x - 1}{3x^2} = \frac{-(\sin x)^2}{(\cos x + 1)3x^2} \to -\frac{1}{6} \text{ per } x \to 0 \ ,$$

utilizzando il fatto che $\sin x/x \to 1$ se $x \to 0$.

Si può anche procedere osservando che anche il rapporto $(\cos x - 1)/3x^2$ presenta la forma di indecisione $0/0$ e ad esso si può riapplicare il teorema ottenendo $-\dfrac{\sin x}{6x} \to -\dfrac{1}{6}$ per $x \to 0$.

In ogni caso si conclude che il limite proposto è $-1/6$.

2.7. Ritroviamo un risultato già noto:

$$\forall a > 1, \forall b > 0 \quad \lim_{x \to +\infty} \frac{x^b}{a^x} = 0 \ ;$$

cioè: *ogni esponenziale con base maggiore di 1 è infinito di ordine superiore rispetto a qualunque potenza di x.*

Scriviamo

$$\frac{x^b}{a^x} = \left(\frac{x}{a^{\frac{x}{b}}}\right)^b = \left(\frac{x}{\alpha^x}\right)^b \quad \text{con } \alpha = a^{\frac{1}{b}} > 1 \ .$$

Essendo le potenze funzioni continue, è sufficiente mostrare che $x/\alpha^x \to 0$ se $x \to +\infty$. Usiamo il teorema 2.6 con $f(x) = x$ e $g(x) = \alpha^x$.

Poiché

$$\lim_{x \to +\infty} \frac{1}{\alpha^x \log \alpha} = 0$$

si conclude che anche

$$\lim_{x\to+\infty} \frac{x}{\alpha^x} = 0 .$$

2.8. $\forall a > 1, \forall b > 0 \quad \lim_{x\to+\infty} \frac{\log_a x}{x^b} = 0.$

Cioè: *ogni potenza di x è infinito di ordine superiore rispetto ai logaritmi.*
È sufficiente applicare il teorema 2.6 con $f(x) = \log_a x$ e $g(x) = x^b$.
Poiché

$$\lim_{x\to+\infty} \frac{f'(x)}{g'(x)} = \lim_{x\to+\infty} \frac{x^{-1}\log_a e}{bx^{b-1}} = \lim_{x\to+\infty} \frac{\log_a e}{bx^b} = 0$$

si ricava l'asserto.
Si deduce subito che anche

$$\lim_{x\to+\infty} \frac{(\log_a x)^\alpha}{x^b} = 0 \quad \forall\alpha > 0 .$$

2.9. Dimostriamo che

$$\forall a > 1, \forall b > 0 \quad \lim_{x\to 0_+} x^b \log_a x = 0 .$$

È una forma di indecisione $0 \cdot \infty$; così il teorema non funziona.
Si può allora scrivere $\log_a x / x^{-b}$ ed applicare il teorema, o, più semplicemente, fare la sostituzione di variabile $t = 1/x$ con la quale il limite diventa

$$\lim_{t\to+\infty} \frac{-\log_a t}{t^b} = 0$$

per l'esercizio precedente.

2.10. Si vogliano calcolare i seguenti limiti:

a) $\lim_{x\to 0_+} x^{10} \cdot e^{\frac{1}{x}}$

b) $\lim_{x\to 0_+} x^x$

c) $\lim_{x\to+\infty} \sqrt[x]{x}$.

a) È una forma di indecisione $0 \cdot \infty$.
Mettendo la funzione nella forma $e^{1/x}/x^{-10}$ (∞/∞) oppure $x^{10}/e^{-1/x}$ $(0/0)$ e applicando il teorema 2.6 non si risolverebbe la forma di indecisione (lo studente è invitato a provare).
Se poniamo $1/x = t$ il limite diventa

$$\lim_{t\to+\infty} \frac{e^t}{t^{10}}$$

che è $+\infty$ rientrando nel tipo dell'esempio 2.7.

b) Si ha $x^x = e^{x\log x}$. Poiché per l'esempio 2.9 $x \log x \to 0$ se $x \to 0_+$ si ha $\lim\limits_{x\to 0_+} x^x = 1$.

c) Si ha $\sqrt[x]{x} = e^{\frac{1}{x}\log x}$. Poiché per l'esempio 2.8 $\log x/x \to 0$ se $x \to +\infty$ si ha $\lim\limits_{x\to+\infty} \sqrt[x]{x} = 1$.

Dimostrazione del teorema di de L'Hôpital.

I. Caso della forma di indecisione 0/0.

a) Sia *L finito.*
Dalla *iii*), fissato $\varepsilon > 0$ esiste t_0 tale che se $t \in (a, t_0)$

$$L - \varepsilon < \frac{f'(t)}{g'(t)} < L + \varepsilon \, . \tag{2.6}$$

Siano ora $a < y < x < t_0$. Nell'intervallo $[y, x]$ f e g verificano le ipotesi del teorema di Cauchy e quindi esiste $c \in (y, x)$ tale che

$$\frac{f(x) - f(y)}{g(x) - g(y)} = \frac{f'(c)}{g'(c)} \, . \tag{2.7}$$

Dalle (2.6) e (2.7) si ricava, essendo $c \in (a, t_0)$:

$$L - \varepsilon < \frac{f(x) - f(y)}{g(x) - g(y)} < L + \varepsilon \, . \tag{2.8}$$

Passiamo ora al limite per $y \to a_+$ nella (2.8). Poiché $f(y) \to 0, g(y) \to 0$ si ricava che, per ogni $x \in (a, t_0)$,

$$L - \varepsilon < f(x)/g(x) < L + \varepsilon \, .$$

e quindi, per la definizione di limite, che $f(x)/g(x) \to L$ per $x \to a_+$.

b) Sia $L = +\infty$ (i casi $L = -\infty$ e $L = \infty$ sono analoghi).
Dalla *iii*), fissato $M > 0$, esiste t_0 tale che se $t \in (a, t_0)$

$$\frac{f'(t)}{g'(t)} > M \, . \tag{2.9}$$

Ragionando come nel caso a), arriviamo anziché alla (2.8) a $\dfrac{f(x) - f(y)}{g(x) - g(y)} > M$ e successivamente, passando al limite per $y \to a_+$, si ottiene $f(x)/g(x) > M$ per ogni $x \in (a, t_0)$. Dunque $f(x)/g(x) \to +\infty$ per $x \to a_+$.

II. Caso della forma di indecisione $+\infty/+\infty$. Gli altri casi sono analoghi.

a) Sia *L finito.*
Procediamo come nel caso I.a): dato $\varepsilon > 0$, sia t_0 tale che valga la (2.6) e di conseguenza se $a < x < y < t_0$, valga anche la (2.8).

Essendo $g'(x) \neq 0$ e $g(x) \to +\infty$ per $x \to a_+$ si ricava che $g(x) > g(y)$. Moltiplichiamo la (2.8) per $[g(x) - g(y)] > 0$ e poi dividiamo per $g(x)$. Si ottiene, isolando $f(x)/g(x)$:

$$(L-\varepsilon)\left[1-\frac{g(y)}{g(x)}\right]+\frac{f(y)}{g(x)}<\frac{f(x)}{g(x)}<(L+\varepsilon)\left[1-\frac{g(y)}{g(x)}\right]+\frac{f(y)}{g(x)}\ . \tag{2.10}$$

Blocchiamo y e passiamo al limite per $x \to a_+$, osservando che $f(y)/g(x)$ e $g(y)/g(x)$ tendono a zero: si deduce

$$L-\varepsilon<\liminf_{x\to a_+}\frac{f(x)}{g(x)}\le\limsup_{x\to a_+}\frac{f(x)}{g(x)}<L+\varepsilon\ .$$

Poiché ε è arbitrario si ha

$$\lim_{x\to a_+}\frac{f(x)}{g(x)}=L\ .$$

b) Sia $L = +\infty$ (i casi $L = -\infty$ e $L = \infty$ sono analoghi).
Si ragiona come nel caso I.b). Fissiamo $M > 0$ e t_0 tali che per ogni $t \in (a, t_0)$, $[f(x) - f(y)] > M[g(x) - g(y)]$.
Dividendo per $g(x)$ otteniamo

$$\frac{f(x)}{g(x)}>M\left[1-\frac{g(y)}{g(x)}\right]+\frac{f(y)}{g(x)}\ .$$

Bloccando y e facendo $x \to a_+$ si ricava

$$\liminf_{x\to a_+}\frac{f(x)}{g(x)}\ge M$$

da cui $f(x)/g(x) \to +\infty$ per $x \to a_+$. □

Per i più curiosi: dalla dimostrazione si ricava che, se $g(x) \to +\infty$ il teorema vale anche senza l'ipotesi che ... $f(x) \to \pm\infty$ e cioè che ci sia effettivamente una forma di indecisione ∞/∞(!).

Esercizi

14. Calcolare $\lim_{x\to+\infty}\sqrt{x^2+1}/x$. Provare ad utilizzare il teorema 2.6.

15. Calcolare i seguenti limiti

a) $\lim_{x\to 0_+}\sqrt[3]{x}(\log x)^{10}$ b) $\lim_{x\to 0_+}\sin x\cdot e^{\frac{1}{x}}$

c) $\lim_{x\to+\infty}(1+e^{-x})^{x^3}$ d) $\lim_{x\to-\infty}2^x(x^3+\log|x|)$.

16. Calcolare $\lim_{n\to+\infty}e^{-n}\left(\sum_{k=1}^{n}k^2\right)$.

17. Si supponga che $f: \mathbb{R} \to \mathbb{R}$ abbia un asintoto obliquo per $x \to +\infty$, di equazione $y = mx + q$. È vero che $\lim_{x\to+\infty} f'(x) = m$?

18. Trovare l'errore:

$$\frac{\sin x - x}{x^3} = \frac{1}{x^2} \cdot \frac{\sin x}{x} - \frac{1}{x^2} .$$

Dunque, poiché $\sin x/x \to 1$ per $x \to 0$, si ha:

$$\lim_{x\to 0} \frac{\sin x - x}{x^3} = \lim_{x\to 0} \left(\frac{1}{x^2} - \frac{1}{x^2} \right) = 0 .$$

2.5 La formula di Taylor

Svilupperemo in questa sezione un aspetto dell'importante problema dell'*approssimazione di una funzione mediante polinomi*. Vi sono vari modi per realizzare tale approssimazione. Quello trattato qui è di carattere locale, almeno nel suo aspetto più elementare, ed è basato su una formula fondamentale: la formula di Taylor.

Cominciamo col considerare una funzione $f: (a,b) \to \mathbb{R}$, derivabile in $x_0 \in (a,b)$. Se $x \in (a,b)$ possiamo scrivere

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + o(x - x_0) \quad \text{per } x \to x_0 . \tag{2.11}$$

Fissiamo l'attenzione sul polinomio, di primo grado in x, a destra nella (2.11):

$$T(x) = f(x_0) + f'(x_0)(x - x_0) .$$

Il suo grafico è la retta tangente al grafico di f nel punto di ascissa x_0. Esso è caratterizzato dalle condizioni seguenti:

$$T(x_0) = f(x_0) \quad \text{e} \quad T'(x_0) = f'(x_0) . \tag{2.12}$$

Si dice che i grafici di f e T o, più semplicemente, che f e T hanno un *contatto del primo ordine* in $x = x_0$.

La (2.11) si può interpretare come una *formula di approssimazione* di f mediante un polinomio di 1° grado che ha con f un contatto del 1° ordine in $x = x_0$.

Il termine $o(x - x_0)$ rappresenta l'errore $E(x)$ che si commette nell'approssimazione:

$$E(x) = f(x) - T(x) .$$

Ad esempio, sia $f(x) = e^x$ e $x_0 = 0$. La (2.11) in questo caso è

$$e^x = 1 + x + o(x) \quad (\text{per } x \to 0) .$$

Il fatto che $e^x - (1 + x) = o(x)$ dice che l'approssimazione è tanto migliore quanto più x è vicino a 0.

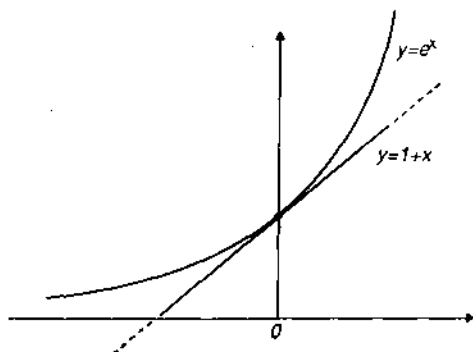



Fig. 6.18

Come si vede dalla figura, allontanandosi un po' da $x = 0$ l'approssimazione di $f(x) = e^x$ tramite $T(x) = 1 + x$ peggiora drasticamente; da qui il carattere locale dell'approssimazione.

Cerchiamo di generalizzare il procedimento, facendo intervenire polinomi di grado più elevato allo scopo di migliorare l'approssimazione.

Definizione 2.3 - *Diremo che due funzioni f e $g: (a,b) \to \mathbb{R}$, hanno un contatto di ordine n in $x_0 \in (a,b)$ se f e g sono derivabili n volte in x_0 e se:*

$$f^{(j)}(x_0) = g^{(j)}(x_0) \quad \forall j = 0, 1, \ldots, n\,; \tag{2.13}$$

cioè se i valori delle funzioni e delle derivate fino all'ordine n incluso coincidono nel punto x_0.

Cerchiamo ora di determinare un polinomio di grado n, $T_n(x)$, che abbia un contatto di ordine n con f in x_0. Successivamente ci porremo il problema di determinare se e in qual senso T_n approssima f.

Guidati dal caso $n = 1$, cerchiamo T_n nella forma seguente:

$$T_n(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + \cdots + a_n(x - x_0)^n$$

e imponiamo a T_n le $(n+1)$ condizioni (2.13) in modo da determinare i coefficienti $a_0, a_1, \ldots, a_n$.

Si ha:

$$\begin{aligned} T_n'(x) =& a_1 + 2a_2(x - x_0) + 3a_3(x - x_0)^2 + \\ & + 4a_4(x - x_0)^3 + \ldots + na_n(x - x_0)^{n-1} \\ T_n''(x) =& 2a_2 + 3 \cdot 2a_3(x - x_0) + \\ & + 4 \cdot 3a_4 \cdot (x - x_0)^2 + \ldots + n(n-1)a_n(x - x_0)^{n-2} \end{aligned}$$

$$T_n'''(x) = 3 \cdot 2a_3 + 4 \cdot 3 \cdot 2a_4(x - x_0) + \ldots + n(n-1)(n-2)a_n(x - x_0)^{n-3}$$

$$\vdots$$

$$T_n^{(n)}(x) = n(n-1) \cdot \ldots \cdot 3 \cdot 2a_n = n!a_n .$$

Quindi:

$$T_n(x_0) = a_0 \ , \ T_n'(x_0) = a_1 \ , \ T_n''(x_0) = 2a_2 \ ,$$

$$T_n'''(x_0) = 3!a_3, \ldots, T_n^{(n)}(x_0) = n!a_n \ .$$

Dalle (2.13):

$$a_0 = f(x_0) \ , \ a_1 = f'(x_0) \ ,$$

$$a_2 = \frac{f''(x_0)}{2} \ , \ a_3 = \frac{f'''(x_0)}{3!} \ , \ \ldots \ , \ a_n = \frac{f^{(n)}(x_0)}{n!} \ .$$

In conclusione abbiamo dimostrato la seguente

Proposizione 2.7 - *Sia $f : (a,b) \to \mathbb{R}$ derivabile n volte in $x_0 \in (a,b)$. Allora esiste un unico polinomio $T_n(x)$ di grado $\leq n$, che abbia un contatto di ordine n con f in x_0, dato dalla formula che segue:*

$$T_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k \ . \tag{2.14}$$

T_n prende il nome di polinomio di Taylor (*) di grado n generato da f, con centro in x_0. Se $x_0 = 0$, T_n assume la forma

$$T_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} x^k \ . \tag{2.15}$$

In questo caso T_n è detto anche polinomio di Mac Laurin di grado n.

***Esempio* 2.11** - Calcoliamo il polinomio di Taylor di ordine 3 della funzione $f(x) = 1/(1+x)$ con centro in $x_0 = 1$.

Abbiamo:

$$f(1) = 1/2$$

$$f'(x) = -(1+x)^{-2} \quad \Rightarrow f'(1) = -\frac{1}{4}$$

$$f''(x) = 2(1+x)^{-3} \quad \Rightarrow f''(1) = \frac{1}{4}$$

$$f'''(x) = -6(1+x)^{-4} \Rightarrow f'''(1) = -\frac{3}{8} \ .$$

(*) Brook Taylor (1685-1731); Colin Mac Laurin (1698-1746).

Dunque

$$T_3(x) = \frac{1}{2} - \frac{1}{4}(x-1) + \frac{1}{8}(x-1)^2 - \frac{1}{16}(x-1)^3 \ .$$

Tra i polinomi di Taylor, quelli di Mac Laurin sono i più usati in pratica. Calcoliamoli per le funzioni e^x, $\sin x$, $\cos x$, $\log(1+x)$.

***Esempio* 2.12** - Sia $f(x) = e^x$.
Poiché $f^{(n)}(x) = e^x \ \forall n \in \mathbb{N}$, si ha $f^{(n)}(0) = 1$, $\forall \ n \in \mathbb{N}$.
Dunque

$$T_n(x) = \sum_{k=0}^{n} \frac{x^k}{k!} \ . \tag{2.16}$$

In particolare

$$T_1(x) = 1 + x \ ,$$

$$T_2(x) = 1 + x + \frac{x^2}{2} \ ,$$

$$T_3(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} \ , \ \dots \ .$$

***Esempio* 2.13** - Sia $f(x) = \sin x$. Si ha:

$f'(x) = \cos x \ , \ f''(x) = -\sin x \ , \ f'''(x) = -\cos x \ ;$

$f^{(iv)}(x) = f(x) \ , \ f^{(v)}(x) = f'(x)$ etc. .

Ne segue che:

$f^{(2k)}(0) = 0$ e $f^{(2k+1)}(0) = (-1)^k$.

Dunque:

$$T_{2n+1}(x) = \sum_{k=0}^{n} (-1)^k \frac{x^{2k+1}}{(2k+1)!} \ . \tag{2.17}$$

In particolare:

$$T_1(x) = x \ ,$$

$$T_3(x) = x - \frac{x^3}{3!} \ ,$$

$$T_5(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!}, \dots \ .$$

Si noti la presenza delle sole potenze dispari che riflette il fatto che $\sin x$ è dispari.

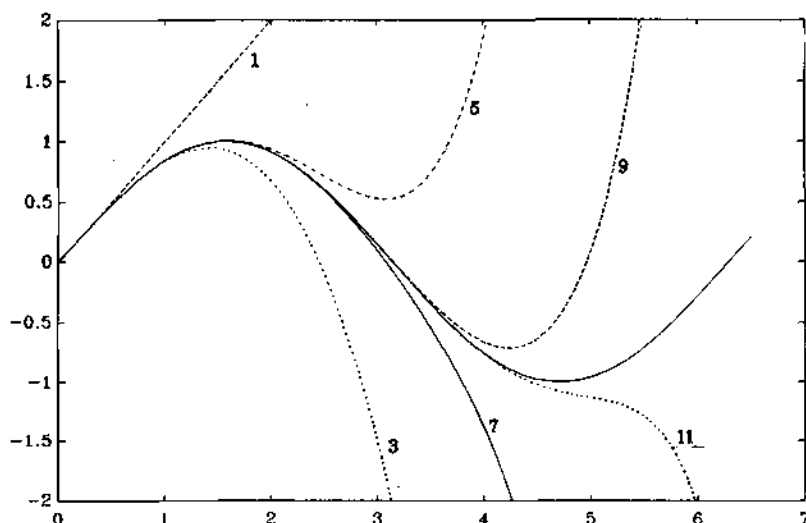



Fig. 6.19 Polinomi di Mac Laurin di grado 1,3,5,7,9,11 per $\sin x$.

***Esempio* 2.14** - Sia $f(x) = \cos x$. Con lo stesso ragionamento usato per $\sin x$ si ottiene:

$$f^{(2k)}(0) = (-1)^k \ , \ f^{(2k+1)}(0) = 0 \ .$$

Dunque:

$$T_{2k}(x) = \sum_{k=0}^{n} (-1)^k \frac{x^{2k}}{(2k)!} \ . \tag{2.18}$$

In particolare:

$$T_2(x) = 1 - \frac{x^2}{2} \ ,$$
$$T_4(x) = 1 - \frac{x^2}{2} + \frac{x^4}{4!} \ ,$$
$$T_6(x) = 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} \ , \ \ldots$$

Si noti la presenza delle sole potenze pari, che riflette il fatto che $\cos x$ è pari.

***Esempio* 2.15** - Sia $f(x) = \log(1 + x)$. Abbiamo:

$$f'(x) = (1 + x)^{-1} \ , \ f''(x) = -(1 + x)^{-2} \ , \ f'''(x) = 2(1 + x)^{-3} \ , \ \ldots \ ,$$
$$f^{(n)}(x) = (-1)^{n+1}(n - 1)!(1 + x)^{-n} \ .$$

Dunque

$$f'(0) = 1 \ , \ f''(0) = -1 \ , \ f'''(0) = 2 \ , \ \dots \ , \ f^{(n)}(0) = (-1)^{n+1}(n-1)! \ .$$

ne segue che

$$T_n(x) = \sum_{k=1}^{n} (-1)^{k+1} \frac{x^k}{k} \ . \tag{2.19}$$

In particolare:

$$T_1(x) = x \ ,$$
$$T_2(x) = x - \frac{x^2}{2} \ ,$$
$$T_3(x) = x - \frac{x^2}{2} + \frac{x^3}{3} \ , \ \dots \ .$$

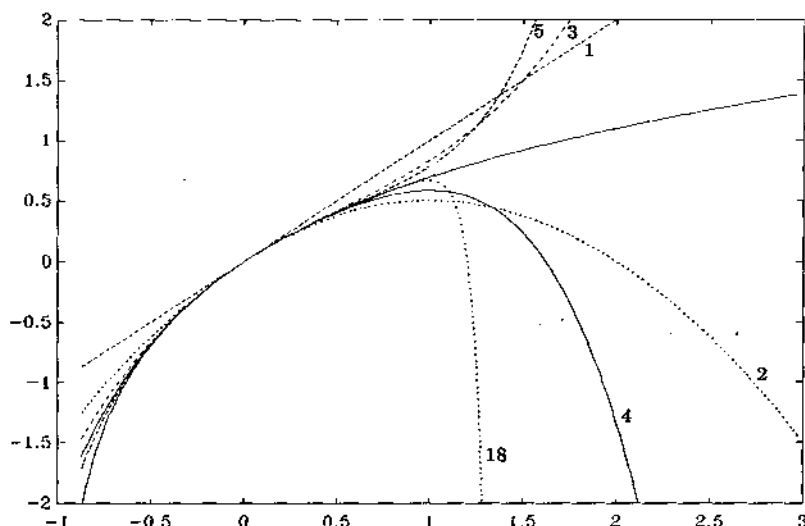



Fig. 6.20 Polinomi di Mac Laurin di grado 1,2,3,4,5,18 per $\log(1 + x)$.

La seguente proposizione mette in luce alcune semplici proprietà dei polinomi di Taylor. Useremo la notazione $T_n[f]$ per indicare il polinomio di Taylor di grado n generato da f.

Proposizione 2.8 - *Valgono le seguenti proprietà. Siano f e g funzioni derivabili n volte in x_0; $c_1, c_2 \in \mathbb{R}$.*

a) $T_n[c_1 f + c_2 g] = c_1 T_n[f] + c_2 T_n[g]$

b) $T_n'[f] = T_{n-1}[f']$.

Lasciamo la facile dimostrazione per esercizio.

Usando la b) si trova ad esempio che il polinomio di Mac Laurin di $f(x) = \frac{1}{1+x}$ (= derivata di $\log(1+x)$), si ottiene per derivazione da quello di $g(x) = \log(1+x)$.

Cioè

$$T_n\left[\frac{1}{1+x}\right] = \sum_{k=0}^{n}(-x)^k .$$

Il problema è ora di valutare *l'errore di approssimazione* o resto n-esimo

$$E_n(x) = f(x) - T_n(x) . \tag{2.20}$$

La (2.20), scritta nella forma

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + E_n(x) \tag{2.21}$$

prende il nome di **formula di Taylor** *di ordine n relativa alla funzione f, con centro in* x_0, (*formula di Mac Laurin* se $x_0 = 0$)

Per avere un'idea del miglioramento che si ottiene aumentando il grado del polinomio approssimante si possono esaminare le seguenti tabelle per la funzione esponenziale.

Approssimazione con un polinomio di grado 2

x	$\exp(x)$	$T_2(x)$	errore
−5.	.006737	8.5	−8.493263
−4.5	.011108	6.625	−6.613892
−4	.018315	5	−4.981695
−3.5	.030197	3.625	−3.594803
−3	.049787	2.5	−2.450213
−2.5	.082085	1.625	−1.542915
−2	.135335	1	−.864665
−1.5	.22313	.625	−.40187
−1	.367879	.5	−.132121
−.5	.60653	.625	−.01847
0	1	1	0
.5	1.648721	1.625	.02372
1	2.718281	2.5	.218281
1.5	4.481689	3.625	.856689
2	7.389056	5	2.389056
2.5	12.18249	6.625	5.557494
3	20.08554	8.5	11.58534

Approssimazione con un polinomio di grado 5

x	$\exp(x)$	$T_5(x)$	errore
5	.006737	−12.33333	12.34007
4.5	.011108	−6.853907	6.865015
−4	.018315	−3.533334	3.551649
−3.5	.030197	−1.645052	1.675249
−3	.049787	−.650001	.699788
−2.5	.082085	−.165365	.24745
−2	.135335	.066666	.068669
−1.5	.22313	.210156	.012974
−1	.367879	366666	.001213
−.5	.60653	.60651	.00002
0	1	1	0
.5	1.648721	1.648698	.000023
1	2.718281	2.716667	.001614
1.5	4.481689	4.461718	.01997
2	7.389056	7.266667	.122389
2.5	12.18249	11.67057	.51192
3	20.08554	18.4	1.685535

Approssimazione con un polinomio di grado 8

x	$\exp(x)$	$T_8(x)$	errore
−5	.006737	3.55184	−3.548447
−4.5	.011108	1.435443	−1.424335
−4	.018315	.530158	−.5118431
−3.5	.030197	.190022	−.159826
−3	.049787	.091294	−.041507
−2.5	.082085	.090462	−.008378
−2	.135335	.136507	−.001173
−1.5	.22313	.223222	−.000092
−1	.367879	.367881	−.000002
−.5	.60653	.60653	.000000
0	1	1	0
.5	1.648721	1.648721	.000000
1	2.718281	2.718279	.000002
1.5	4.481689	4.481564	.000124
2	7.389056	7.387302	.001754
2.5	12.18249	12.1686	.01389
3	20.08554	20.00915	.076381

Il prossimo teorema presenta due possibili formule per $E_n(x)$.

▪ **Teorema 2.9** - *Sia $f:(a,b) \to \mathbb{R}$, derivabile n volte in $x_0 \in (a,b)$. Allora se $T_n(x)$ è il polinomio di Taylor di grado n generato da f con centro in x_0, posto $E_n(x) = f(x) - T_n(x)$, si ha:*

a) *Per ogni* $n \in \mathbb{N}$,

$$E_n(x) = o((x - x_0)^n) \quad \textit{per } x \to x_0 \text{ (formula di Peano)} . \tag{2.22}$$

b) *Se f è inoltre derivabile $(n+1)$ volte in (a,b), escluso al più x_0, per ogni $x \in (a,b)$ esiste c compreso tra x_0 ed x, tale che:*

$$E_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x - x_0)^{n+1} \text{ (formula di Lagrange)} . \tag{2.23}$$

Si noti che, fissati f ed x_0, c dipende da x e da n.

La formula di Peano è particolarmente utile nel calcolo dei limiti, come vedremo nella prossima sezione. Si noti che nel caso $n = 1$ la (2.22) non è altro che $f(x) = f(x_0) + f'(x_0)(x - x_0) + o((x - x_0))$, equivalente alla differenziabilità di f in x_0.

La formula di Lagrange si rivela particolarmente utile nel calcolo numerico delle funzioni irrazionali o trascendenti. Un cenno a tale applicazione si trova nella prossima sezione. Il caso $n = 0$ non è altro che il teorema del valor medio di Lagrange (formula $(2.3)'$).

Dimostrazione del teorema 2.9 -

a) Bisogna dimostrare che $\{f(x) - T_n(x)\}/(x - x_0)^n \to 0$ se $x \to x_0$. Usiamo il principio di induzione su n.

Se $n = 1$ la (2.22) è vera per la definizione di derivabilità di f in x_0. Supponiamo che la (2.22) sia vera per l'intero $n - 1$. Ricordiamo che, per la b) della proposizione 2.8, $T_n'(x)$ è il polinomio di Taylor di grado $n - 1$ per f'. L'ipotesi di ricorrenza implica allora che

$$f'(x) - T_n'(x) = o((x - x_0)^{n-1}) . \tag{2.24}$$

Applichiamo la regola di de L'Hôpital al rapporto $[f(x) - T_n(x)]/(x - x_0)^n$. Si ottiene, in base alla (2.24)

$$\lim_{x \to x_0} \frac{f(x) - T_n(x)}{(x - x_0)^n} = \lim_{x \to x_0} \frac{f'(x) - T_n'(x)}{n(x - x_0)^{n-1}} = 0 .$$

b) Sia $x > x_0$. Il caso $x < x_0$ è analogo. Poniamo $g(x) = f(x) - T_n(x)$ e $w(x) = (x - x_0)^{n+1}$. Osserviamo che, essendo T_n il polinomio di Taylor di f:

$$g(x_0) = g'(x_0) = g''(x_0) = \cdots = g^{(n)}(x_0) = 0 \text{ e } g^{(n+1)}(x) = f^{(n+1)}(x) \tag{2.25}$$

per ogni $x \in (a,b)$.

Inoltre:

$$w(x_0) = w'(x_0) = \cdots = w^{(n)}(x_0) = 0 \quad \text{e} \quad w^{(n+1)}(x) = (n+1)! . \tag{2.26}$$

Nell'intervallo $[x_0, x]$ le funzioni g e w soddisfano le ipotesi del teorema di Cauchy; ne segue che esiste un punto $c_1 \in (x_0, x)$ tale che

$$\frac{g(x)}{w(x)} = \frac{g'(c_1)}{w'(c_1)} .$$

Le funzioni g' e w' soddisfano le ipotesi del teorema di Cauchy in $[x_0, c_1]$. Dunque, ricordando le (2.25) e (2.26), esiste un punto $c_2 \in (x_0, c_1)$ tale che

$$\frac{g'(c_1)}{w'(c_1)} = \frac{g''(c_2)}{w''(c_2)} .$$

Si procede in questo modo fino ad ottenere una sequenza di $(n+1)$ numeri $c_1, c_2, \ldots, c_{n+1}$ tali che

$$x_0 < c_{n+1} < c_n < \ldots < c_2 < c_1$$

e che

$$\frac{g(x)}{w(x)} = \frac{g'(c_1)}{w'(c_1)} = \frac{g''(c_2)}{w''(c_2)} = \cdots = \frac{g^{(n+1)}(c_{n+1})}{w^{(n+1)}(c_{n+1})} .$$

Ricordando che $g^{(n+1)} = f^{(n+1)}$ e $w^{(n+1)} = (n+1)!$, sostituendo a g e w le loro espressioni si ricava

$$\frac{f(x) - T_n(x)}{(x - x_0)^{n+1}} = \frac{f^{(n+1)}(c_{n+1})}{(n+1)!}$$

che coincide con la tesi del teorema ponendo $c = c_{n+1}$. □

Ecco le formule di Mac Laurin con il resto di Peano per le funzioni e^x, $\cos x$, $\sin x$, $\log(1+x)$; si intende che gli "o" sono tutti per $x \to 0$.

***Esempio* 2.16 -**

a. $e^x = \sum_{k=0}^{n} \frac{x^k}{k!} + o(x^n)$

b. $\sin x = \sum_{k=0}^{n} (-1)^k \frac{x^{2k+1}}{(2k+1)!} + o(x^{2n+2})$

c. $\cos x = \sum_{k=0}^{n} (-1)^k \frac{x^{2k}}{(2k)!} + o(x^{2n+1})$

d. $\log(1+x) = \sum_{k=1}^{n} (-1)^{k+1} \frac{x^k}{k} + o(x^n)$.

Conoscendo la formula di Taylor per una funzione, se ne possono costruire rapidamente altre usando la proposizione 2.8 e la seguente.

Proposizione 2.10 - *Sia $f: (a,b) \to \mathbb{R}$, derivabile n volte in $x_0 = 0 \in (a,b)$. Se esiste un polinomio P_n di grado $\leq n$, tale che*

$$f(x) = P_n(x) + o(x^n) \quad \textit{per } x \to 0 , \tag{2.27}$$

allora $P_n = T_n[f]$, polinomio di Mac Laurin per f.

Dimostrazione - Osserviamo che se w è derivabile n volte in 0 e $w(0) = w'(0) = \cdots = w^{(n-1)}(0) = 0$ allora, applicando $(k-1)$ volte la regola di de L'Hôpital si trova per $k = 1, \ldots, n$:

$$w^{(k)}(0) = \lim_{x\to 0} \frac{w(x)}{x^k} k! \,. \tag{2.28}$$

Ponendo $w(x) = f(x) - P_n(x)$, dalle (2.27), (2.28) si ha per $k = 1, 2, \ldots, n$:

$$w^{(k)}(0) = \lim_{x\to 0} \frac{w(x)}{x^k} k! = \lim_{x\to 0} k! \frac{o(x^n)}{x^k} = 0 \,.$$

Essendo poi $w(0) = 0$ si ottiene che $f^{(k)}(0) = P_n^{(k)}(0)\ \forall k = 0, \ldots, n$.
Cioè $P_n(x) = T_n[f](x)$. □

La proposizione 2.10 legittima in ultima analisi il procedimento di sostituzione nei prossimi esempi.

***Esempio* 2.17** - Scriviamo la formula di Mac Laurin per le funzioni:

$$e^{-x} \ , \ \sin x^2 \ , \ \log(1 - x^3) \,.$$

Sostituendo $(-x), x^2, -x^3$ ad x nelle formule a, b., d. dell'esempio 2.16 rispettivamente si ottiene:

a. $e^{-x} = 1 - x + \dfrac{x^2}{2} - \dfrac{x^3}{3!} + \cdots + (-1)^n \dfrac{x^n}{n!} + o(x^n)$

b. $\sin x^2 = x^2 - \dfrac{x^6}{3!} + \dfrac{x^{10}}{5!} + \cdots + (-1)^n \dfrac{x^{4n+2}}{(2n+1)!} + o(x^{4n+4})$

c. $\log(1 - x^3) = -x^3 - \dfrac{x^6}{2} - \dfrac{x^9}{3} - \cdots - \dfrac{x^{3n}}{n} + o(x^{3n})$

(dove gli "o" sono sempre da intendersi per $x \to 0$).
Concludiamo il paragrafo con le formule di Mac Laurin (resto di Peano) per altre funzioni che intervengono sovente nei calcoli, lasciando allo studente la loro verifica.

***Esempio* 2.18** -

a. $\operatorname{SettTh} x = x + \dfrac{x^3}{3} + \dfrac{x^5}{5} + \cdots + \dfrac{x^{2n+1}}{2n+1} + o(x^{2n+2})$

b. $\operatorname{arctg} x = x - \dfrac{x^3}{3} + \dfrac{x^5}{5} + \cdots + (-1)^n \dfrac{x^{2n+1}}{2n+1} + o(x^{2n+2})$

c. $\operatorname{Sh} x = x + \dfrac{x^3}{3!} + \dfrac{x^5}{5!} + \cdots + \dfrac{x^{2n+1}}{(2n+1)!} + o(x^{2n+2})$

d. $\operatorname{Ch} x = 1 + \frac{x^2}{2} + \frac{x^4}{4!} + \cdots + \frac{x^{2n}}{(2n)!} + o(x^{2n+1})$

e. $(1+x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2}x^2 + \frac{\alpha(\alpha-1)(\alpha-2)}{3!}x^3 + \cdots +$
$$+ \frac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!}x^n + o(x^n)$$

dove $\alpha \in \mathbb{R}$; si usa porre, in analogia con la notazione già introdotta quando $\alpha \in \mathbb{N}$:

$$\binom{\alpha}{n} := \frac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!} \qquad \alpha \in \mathbb{R}, n \in \mathbb{N} . \tag{2.29}$$

Si osserva che, se $\alpha \notin \mathbb{N}, \binom{\alpha}{n}$ è sempre diverso da 0 per ogni $n \in \mathbb{N}$.
Risulta, per esempio,

$$\binom{-1}{n} = (-1)^n \ , \ \left\{\binom{\frac{1}{2}}{n}\right\}_{n=0,1,2,\dots} = \left\{1, \frac{1}{2}, -\frac{1}{8}, \frac{1}{16}, \dots\right\} .$$

da cui segue

$$\sqrt{1+x} = 1 + \frac{x}{2} - \frac{x^2}{8} + \frac{3}{16}x^3 + o(x^3) .$$

Esercizi

19. Dimostrare le formule dell'esempio 2.18 a, b, c, d, e.

20. Sia $f:(a,b) \to \mathbb{R}$ tale che: f è due volte derivabile in (a,b); nulla per un $x_0 \in (a,b)$ e $f''(x) + f(x) = 0$, $\forall x \in (a,b)$. Dimostrare che $f = 0$.

21. È vero che il polinomio di Mac Laurin $T_3[e^{x^3+2}]$ è: $e^2 + e^2 x^3$?

22. Scrivere le formule di Mac Laurin arrestate all'ordine n per le funzioni indicate, con il resto di Peano:

a) $f(x) = e^{x^3}$ b) $f(x) = \sin\left(\frac{x}{2}\right)$ c) $f(x) = (1-x)^{\frac{1}{3}}$

d) $f(x) = \operatorname{arctg} x^2$ e) $f(x) = \operatorname{Ch}(2x)$ f) $f(x) = \log(1-x) + \sin x$.

23. È vero che se $f(x) = o(x^2)$ per $x \to 0$ allora $f(0) = f'(0) = f''(0) = 0$? E se aggiungiamo l'ipotesi $f \in C^2(-1,1)$?

24. Per le seguenti funzioni trovare $P_n(x)$, polinomio di grado n, tale che $f(x) - P_n(x) = o(x^n)$. È vero che $P_n = T_n[f]$?

a) $f(x) = \frac{\sin x}{x}$ b) $f(x) = \frac{\sin x - x}{x^3}$ c) $f(x) = \frac{1}{x}(\cos x - e^x)$.

25.* Sia $f: \mathbb{R} \to \mathbb{R}$, positiva e derivabile 2 volte con $f''(z) \leq M$, $M > 0$, $\forall z \in \mathbb{R}$. Dimostrare che, $\forall x \in \mathbb{R}$

$$|f'(x)| < \sqrt{2Mf(x)}\,.$$

[Suggerimento: scrivere la formula di Taylor con centro x, arrestata al 2° ordine ed usare l'ipotesi su f''.]

26. Scrivere la formula di Mac Laurin arrestata al 3° ordine per le funzioni $\sin x$ ed e^x con il resto di Lagrange.

27. Dopo aver verificato che la funzione seguente è differenziabile infinite volte, scrivere il polinomio di Mac Laurin di grado n

$$f(x) = \begin{cases} e^{-\frac{1}{x}} & x > 0 \\ 0 & x \leq 0\,. \end{cases}$$

28. Scrivere il polinomio di Mac Laurin di grado n per le funzioni: arcsin e arccos.

3. ALCUNE APPLICAZIONI

3.1 Funzioni convesse e concave

Una importante classe di funzioni è quella delle funzioni convesse su un intervallo $I \subseteq \mathbb{R}$.

La condizione di convessità di una funzione $f: I \to \mathbb{R}$ può essere espressa geometricamente in maniera molto semplice facendo intervenire l'*epigrafico* di f, $\mathrm{Epi}(f)$, cioè la parte di piano che sta al di sopra o sul grafico di f. In formule:

$$\mathrm{Epi}(f) := \{(x, y) \in \mathbb{R}^2 : x \in I \quad \text{e} \quad y \geq f(x)\}\,.$$

Fig. 6.21

a) Funzione convessa. *b)* Funzione concava.

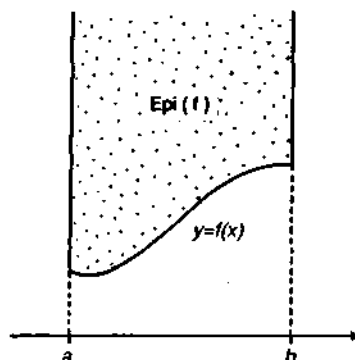
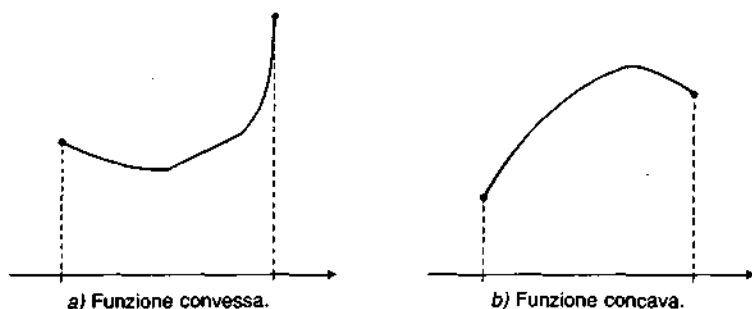



Fig. 6.22

Definizione 3.1 - *$f: I \to \mathbb{R}$ si dice convessa in I se* $\mathrm{Epi}(f)$ *è un insieme convesso; f si dice concava se $-f$ è convessa.*

Ricordando la definizione di insieme convesso data in 3.2.4 è facile verificare che la definizione 3.1 è equivalente alla seguente:

Definizione 3.1' - *$f: I \to \mathbb{R}$ è convessa (concava) su I se per ogni coppia di punti $x, y \in I$ il segmento di estremi $(x, f(x))$ e $(y, f(y))$ non ha punti sotto (sopra) il grafico di f.*

Le funzioni convesse (concave) possono avere tratti del grafico coincidenti con segmenti di retta (vedi fig. 6.22a). Le funzioni convesse (concave) che *non* hanno tratti rettilinei del grafico si dicono *strettamente* convesse (concave). In questo caso, $\forall x, y \in I$, il segmento di estremi $(x, f(x))$, $(y, f(y))$ sta tutto *sopra (sotto)* il grafico di f tranne ovviamente agli estremi.

La condizione di convessità 3.1' si può esprimere analiticamente nel seguente modo, in riferimento alla figura 6.23.

Esprimiamo il generico punto (z_t, f_t) del segmento congiungente i punti $(x, f(x))$, $(y, f(y))$ come *combinazione lineare convessa* dei suoi estremi (cfr. 3.2.5)

$$z_t = (1-t)x + ty \ , \ \ f_t = (1-t)f(x) + tf(y) \ , \ \ t \in [0,1]$$

Fig. 6.23

Funzione convessa con due punti angolosi

Funzione convessa con due discontinuità agli estremi di un intervallo

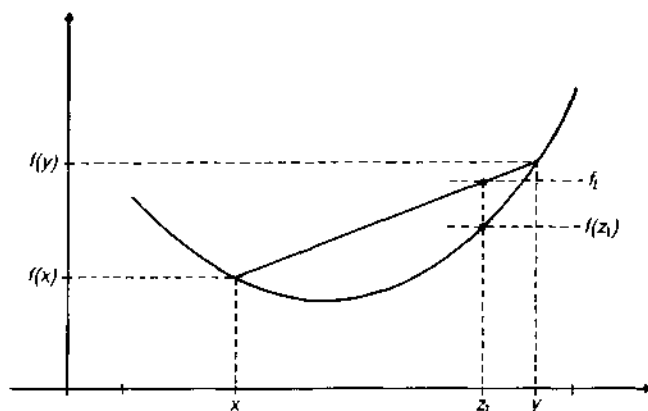
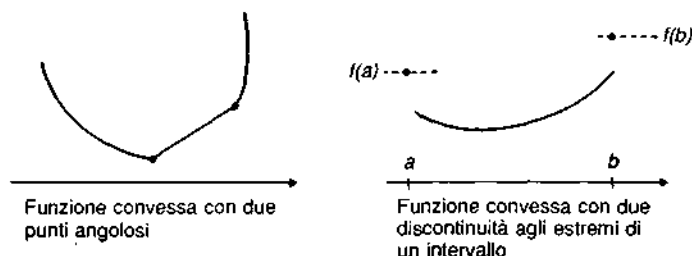



Fig. 6.24

La definizione 3.1′ si esprime allora analiticamente così: $\forall x, y \in I$, $\forall t \in (0,1)$

$$f(z_t) \leq f_t \ . \tag{3.1}$$

Se nella (3.1) vale $<$, f è *strettamente* convessa.

Per le funzioni concave si ha $\forall x, y \in I$, $\forall t \in (0,1)$, $f(z_t) \geq f_t$ con il $>$ per le strettamente concave.

Le funzioni convesse e concave possono presentare punti angolosi e discontinuità, ma solo di prima specie e agli estremi dell'intervallo (fig. 6.24).

Se ci limitiamo ad intervalli $I = (a, b)$, aperti, vale il seguente teorema

■ **Teorema 3.1** - *Sia* $f: (a,b) \to \mathbb{R}$, *convessa o concava. Allora*

a) $\forall x \in (a,b)$ *esistono* $f'_+(x)$ *ed* $f'_-(x)$

b) f *è continua in* (a,b).

Dimostrazione - Si basa sul seguente lemma che caratterizza la convessità (concavità) in base alle pendenze dei segmenti che congiungono due punti del grafico di una funzione.

Per $f: I \to \mathbb{R}$, conviene porre $\Phi(z, x) = \dfrac{f(z) - f(x)}{z - x}$, rapporto che rappresenta la pendenza del segmento congiungente $(x, f(x))$ e $(z, f(z))$. Ci limitiamo a funzioni convesse.

Lemma 3.2 - $f: I \to \mathbb{R}$ è *convessa se e solo se*

$$\forall x, y, z \in I \ , \ x < z < y \Rightarrow \Phi(z, x) \leq \Phi(y, x) \leq \Phi(y, z) \ . \tag{3.2}$$

Se f è strettamente convessa i segni di $\leq$ vanno sostituiti con $<$.

Omettiamo la dimostrazione del lemma, limitandoci solo ad osservare che esprime un fatto "quasi ovvio" (Fig. 6.25).

La (3.2) indica che la pendenza del segmento AC è intermedia tra quelle dei segmenti AB e BC. In particolare si deduce che f è convessa se e solo se i rapporti incrementali $\Phi(y, x)$ sono crescenti in y per x fissato.

La dimostrazione del teorema 3.1 è ora immediata: i limiti $\lim\limits_{y \to x_\pm} \Phi(y, x) = f'_\pm(x)$, $\forall x \in (a, b)$, esistono finiti in base al teorema di esistenza del limite delle funzioni monotone; ciò implica continuità dalla destra e continuità dalla sinistra per f; dunque f è continua. □

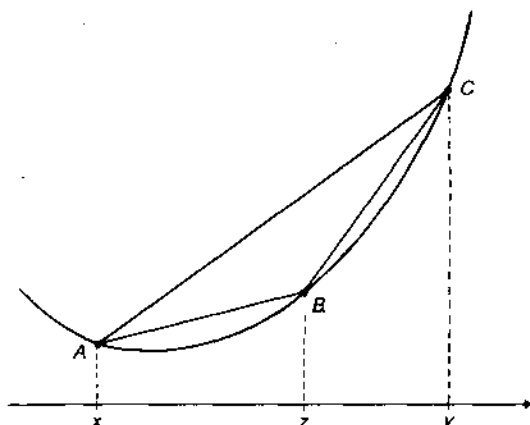



Fig. 6.25

Sia $f\colon(a,b) \to \mathbb{R}$ *strettamente convessa* e derivabile in un punto $x_0 \in (a,b)$; utilizzando sempre il lemma 3.2 otteniamo che:

$$\Phi(x,x_0) < \Phi(z,x_0) \quad \text{se} \quad x < z < x_0$$

$$\Phi(x,x_0) > \Phi(z,x_0) \quad \text{se} \quad x > z > x_0 \ .$$

Dunque, nel primo caso

$$\Phi(x,x_0) < \lim_{z \to x_0} \Phi(z,x_0) = f'(x_0)$$

e nel secondo

$$\Phi(x,x_0) > \lim_{z \to x_0} \Phi(z,x_0) = f'(x_0) \ .$$

In entrambi i casi si ricava:

$$f(x) > f(x_0) + f'(x_0)\,(x - x_0) \quad \text{per } x \neq x_0 \ , \tag{3.3}$$

ovvero: *se f è convessa e derivabile in x_0, la retta tangente al grafico di f in $(x_0, f(x_0))$ sta sotto il grafico di f in tutto (a,b), tranne che in x_0.*

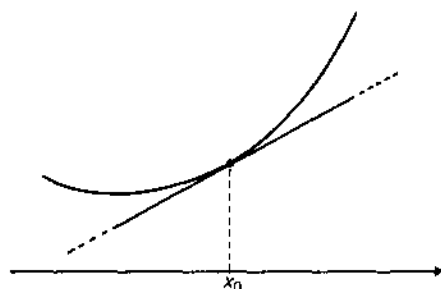



Fig. 6.26

Osservazione 3.1 - La (3.3) implica che: *se x_0 è stazionario per f strettamente convessa in (a,b), allora $f(x) > f(x_0)$, $\forall x \in (a,b)$, $x \neq x_0$.*

Ne segue che

$$f(x_0) = \min_{x \in (a,b)} f(x) .$$

È questa proprietà che rende estremamente utili le funzioni convesse nella teoria dell'ottimizzazione.

Naturalmente le stesse considerazioni, con ovvi cambiamenti valgono per le funzioni concave. Ad esempio la (3.3) vale col $<$ per le funzioni strettamente concave.

Il prossimo teorema è molto usato per riconoscere la convessità o concavità di una funzione.

■ **Teorema 3.3** - *Sia $f:(a,b) \to \mathbb{R}$.*

a) *se f è derivabile in (a,b) f è convessa (concava) se e solo se f' è crescente (decrescente)*

b) *se f è derivabile due volte in (a,b) f è convessa (concava) se e solo se $f''(x) \geq 0$ ($f''(x) \leq 0$) $\forall x \in (a,b)$.*

Il teorema si modifica in maniera ovvia per le funzioni strettamente convesse o concave: nella a) "crescente" (decrescente) è sostituito da "strettamente crescente" (strettamente decrescente); nella b) $f''(x) \geq 0$ ($f''(x) \leq 0$) è sostituito con $f''(x) > 0$ ($f''(x) < 0$); "se e solo se" è sostituito con "se".

Dimostrazione - a) Dalla (3.2), se $x < z < y$ abbiamo

$$f'(x) = \lim_{z \to x} \Phi(z,x) \leq \Phi(y,x) .$$

D'altra parte

$$\Phi(y,x) \leq \lim_{z \to y} \Phi(y,z) = f'(y) .$$

Dunque $f'(x) \leq f'(y)$ ed f' è crescente.

Viceversa sia f' crescente.

Per ogni terna $x, y, z \in (a,b)$, $x < z < y$, dal teorema del valor medio, esistono c e d in (a,b) tali che

$$\Phi(z,x) = f'(c) \quad \text{con } x < c < z$$

e

$$\Phi(y,z) = f'(d) \quad \text{con } z < d < y .$$

Essendo $c < d$ si ha $f'(d) \geq f'(c)$ da cui

$$\Phi(x,z) = \Phi(z,x) \leq \Phi(y,z) .$$

Dunque $\Phi(x,z)$ è crescente in x, con z fissato e perciò f è convessa. b) segue subito dalla a) poiché f' è crescente se e solo se $f'' \geq 0$. □

Esempi

3.1. Sia $f(x) = a^x$, $a > 0$; essendo $f''(x) = (\log a)^2 a^x > 0$, a^x è strettamente convessa in $\mathbb{R}$.

3.2. Sia $f(x) = \log_a x$, $a \neq 1$, $x \in \mathbb{R}_+$.
Si ha:

$$f''(x) = -\frac{1}{x^2 \log a} .$$

Dunque se $0 < a < 1$, $f''(x) > 0$ $\forall x \in \mathbb{R}_+$ e $\log_a x$ è strettamente convessa. Se $a > 1$, $f''(x) < 0$, cioè $\log_a x$ è strettamente concava.

3.3. $f(x) = x^\alpha$, $x \in \mathbb{R}_+$.
Si ha: $f''(x) = \alpha(\alpha - 1)x^{\alpha-2}$. Se $\alpha > 1$, $f''(x) > 0$ $\forall x \in \mathbb{R}_+$ e quindi x^α è strettamente convessa. Se $\alpha = 0$ o $\alpha = 1$, $f(x)$ ha come grafico una retta: concava e convessa nello stesso tempo.
Se $0 < \alpha < 1$, $f''(x) < 0$ e quindi x^α è strettamente concava in $\mathbb{R}_+$. Se infine $\alpha < 0$, $f''(x) > 0$; dunque x^α è strettamente convessa in $\mathbb{R}_+$.

Veniamo ora alla definizione dei punti di *flesso*.

Definizione 3.2 - *Sia $f: (a,b) \to \mathbb{R}$ ed $x_0 \in (a,b)$ un punto di derivabilità per f, oppure $f'(x_0) = \pm\infty$. Il punto x_0 si dice di* **flesso** *se esiste un intorno destro di x_0 in cui f è convessa (concava) ed un intorno sinistro di x_0 in cui f è concava (convessa).*

Evidentemente se x_0 è di flesso per f, la tangente nel punto $(x_0, f(x_0))$ "attraversa" il grafico di f: vedi Fig. 6.27. Non è vero il viceversa come mostra l'esercizio 2.

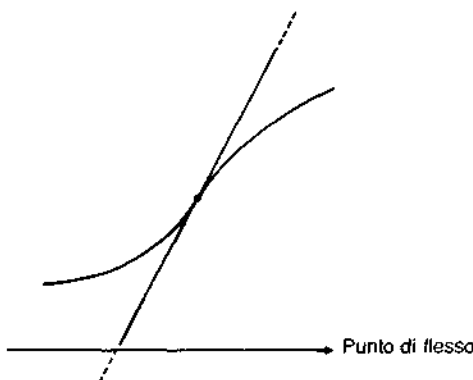



Fig. 6.27

Per funzioni due volte derivabili i punti di flesso si trovano tra quelli che annullano la derivata seconda:

Proposizione 3.4 - *Sia x_0 punto di flesso per f; se esiste $f''(x_0)$, allora deve essere nulla.*

Naturalmente vi sono punti a derivata seconda nulla che *non* sono di flesso. Ad esempio $f(x) = x^4$ ha $f''(0) = 0$ ma $x = 0$ è punto di minimo.

Esercizi

1. Dimostrare che se $f: (a, b) \to \mathbb{R}$ è convessa allora f'_+ ed f'_- sono crescenti. Inoltre $\forall x \in (a,b)$, $f'_-(x) \le f'_+(x)$.

2. Sia

$$f(x) = \begin{cases} x^2 \sin \frac{1}{x} & x \neq 0 \\ 0 & x = 0 \end{cases} .$$

Dimostrare che $x = 0$ è punto stazionario per f ma che non è di estremo né di flesso.

3. Determinare zone di concavità-convessità ed eventuali punti di flesso per le seguenti funzioni:

$$\text{a) } f(x) = \sqrt{(x-1)(x-2)} \quad \text{b) } f(x) = \frac{\log x}{x} \quad \text{c) } f(x) = xe^{-x^2} .$$

4.* Sia $f: (a, b) \to \mathbb{R}$, convessa. Siano $p_1, p_2, \ldots, p_n$ numeri reali *non negativi* tali che $\sum_{j=1}^{n} p_j = 1$.

Dimostrare che, $\forall x_1, x_2, \ldots, x_n \in (a, b)$, si ha:

$$f\left(\sum_{j=1}^{n} p_j x_j\right) \le \sum_{j=1}^{n} p_j f(x_j) \qquad \textit{(disuguaglianza di Jensen)} .$$

[Suggerimento: usare il principio di induzione a partire da $n = 2$. Osservare poi che $\sum_{j=1}^{n} p_j x_j = (1 - p_n) \sum_{j=1}^{n-1} \frac{p_j}{1-p_n} x_j + p_n x_n$ e che $\sum_{j=1}^{n-1} \frac{p_j}{1-p_n} = 1 \ldots$].

5. Sia $f(a, b) \to \mathbb{R}$ convessa e $x_1, x_2, \ldots, x_n \in (a, b)$.
Dimostrare che

$$f\left(\frac{1}{n}\sum_{j=1}^{n} x_j\right) \le \frac{1}{n}\sum_{j=1}^{n} f(x_j) .$$

(utilizzare l'esercizio 4.*).

6.* Verificare che $f(x) = x \log x$ è convessa in $\mathbb{R}_+$.

Utilizzando l'esercizio 5. verificare che, se $x_1, x_2, \ldots, x_n \in \mathbb{R}_+$ e $\sum_{j=1}^{n} x_j = 1$, la funzione

$$H(x) := -\sum_{j=1}^{n} x_j \log x_j$$

è massima quando $x_1 = x_2 = \ldots = x_n = 1/n$.

Se si interpretano i numeri x_j come probabilità di un evento E_j (infatti $0 < x_j \leq 1$) la funzione H si chiama **entropia** del sistema $\{E_j, x_j\}_{j=1,\ldots,n}$.

7.* a) Verificare che, se $p > 1$ e $q = \dfrac{p}{p-1}$ (cioè $1/p + 1/q = 1$), per ogni $u > 0$ si ha:

$$u^{\frac{1}{p}} < \frac{u}{p} + \frac{1}{q} \tag{3.4}$$

(usare la stretta concavità di u^α per $\alpha < 1$ e usare la (3.3) col "<").

b) Porre nella (3.4) $u = x^p y^{-q}$, $x, y \in \mathbb{R}_+$ e dedurre la disuguaglianza

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q} \qquad \textit{(disuguaglianza di Young)} .$$

8. Sia $\Psi = g \circ f$ con f convessa e g convessa crescente. Dimostrare che Ψ è convessa. (Ad esempio f convessa $\Rightarrow e^f$ convessa).

9.* Dimostrare che se f è derivabile in (a, b), f è convessa se e solo se

$$\forall x, x_0 \in (a, b) \ , \ f(x) \geq f(x_0) + f'(x_0)(x - x_0) .$$

Interpretare geometricamente il risultato.

3.2 Applicazioni della formula di Taylor

Tra le più importanti applicazioni della formula di Taylor annoveriamo le seguenti:

A. Determinazione della natura dei punti stazionari

B. Calcolo di ordini di infinitesimo e di limiti

C. Calcolo numerico.

A. Supponiano che una funzione $f:(a, b) \to \mathbb{R}$ abbia un punto stazionario in $x_0 \in (a, b)$. Per determinarne la natura, se cioè esso è o non è un punto di estremo locale occorre valutare il segno dell'incremento

$$f(x) - f(x_0)$$

al variare di x in un intorno di x_0. Più precisamente, se $f(x) - f(x_0) \geq 0$ (risp. ≤ 0) in un intorno di x_0, allora x_0 è punto di minimo (risp. massimo) locale. Se al contrario, in *ogni* intorno di x_0, $f(x) - f(x_0)$ cambia segno, x_0 non può essere punto di estremo.

Il seguente teorema indica come si può valutare il segno di $f(x) - f(x_0)$ usando la formula di Taylor.

■ **Teorema 3.5** - *Sia $f:(a, b) \to \mathbb{R}$ derivabile n volte in $x_0 \in (a, b)$, $n \geq 2$, e tale che in x_0 tutte le derivate tranne l'n-esima siano nulle:*

$$f'(x_0) = f''(x_0) = \cdots = f^{(n-1)}(x_0) = 0 \ , \ f^{(n)}(x_0) \neq 0 .$$

Allora abbiamo le seguenti alternative:

$$n \text{ pari} \begin{cases} f^{(n)}(x_0) > 0 & \Rightarrow \quad x_0 \text{ è punto di minimo locale forte;} \\ \\ f^{(n)}(x_0) < 0 & \Rightarrow \quad x_0 \text{ è punto di massimo locale forte;} \end{cases}$$

$$n \text{ dispari} \qquad\qquad \Rightarrow \quad x_0 \text{ non è punto di estremo.}$$

Dimostrazione - Usiamo la formula di Taylor, arrestata all'ordine n con resto di Peano; poiché l'unica derivata non nulla in x_0 è l'ultima si trova:

$$f(x) - f(x_0) = \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + o((x - x_0)^n)$$
$$= (x - x_0)^n \left[\frac{f^{(n)}(x_0)}{n!} + o(1)\right] \quad (\text{per } x \to x_0) .$$

Sia n *pari*. Allora $(x - x_0)^n$ è sempre positivo per $x \neq x_0$; dunque il segno dell'incremento $f(x) - f(x_0)$ è determinato dal fattore entro parentesi quadre. Poiché $o(1) \to 0$ per $x \to x_0$, il segno di tale fattore sarà definitivamente quello di $f^{(n)}(x_0)$.

Quindi, se $f^{(n)}(x_0) > 0$ (risp. < 0) avremo $f(x) - f(x_0) > 0$ (risp. < 0) in un intorno di x_0 e cioè x_0 è punto di minimo (risp. massimo) locale forte.

Sia n *dispari*. Allora $(x - x_0)^n$ cambia segno con $x - x_0$, mentre il fattore tra parentesi quadre ha, per il discorso fatto prima, definitivamente il segno di $f^{(n)}(x_0)$. Quindi $f(x)- -f(x_0)$ cambia segno in *ogni* intorno di x_0 ed x_0 *non* può essere punto di estremo. □

***Esempio* 3.4** - Sia $f(x) = x \sin x - \cos 2x$.

La funzione f è derivabile infinite volte in $\mathbb{R}$. Poiché $f'(x) = \sin x + x \cos x+ +2 \sin 2x$, l'origine è un punto stazionario ($f'(0) = 0$).

Calcoliamo la prima derivata non nulla in 0:

$$f''(x) = 2 \cos x - x \sin x + 4 \cos 2x \ ;$$

poiché $f''(0) = 6 > 0$, concludiamo che $x = 0$ è punto di minimo locale forte per f.

B. La formula di Taylor è una "radiografia" del comportamento di una funzione nell'intorno del centro x_0; per questo motivo è uno strumento efficace nella determinazione dell'ordine di infinitesimo della funzione stessa per $x \to x_0$, rispetto all'infinitesimo campione $(x - x_0)$. La cosa migliore è spiegarsi con esempi.

***Esempio* 3.5** - Si voglia determinare l'ordine di infinitesimo per $x \to 0$ delle seguenti funzioni

a) $\sin x - x$ c) $x \log(1 - x) + e^{x^2} - 1$

b) $(\sin x)^2 - x^2$ d) $\cos 2\sqrt{x} - e^{-2x}$.

La strategia consiste nell'usare la formula di Mac Laurin per arrivare ad una relazione del tipo

$$f(x) = kx^{\alpha} + o(x^{\beta}) \qquad (3.5)$$

con $k \in \mathbb{R}$, $k \neq 0$ e dove (è essenziale!) $\beta \geq \alpha$. L'ordine cercato sarà dunque α.

La (3.5) determina l'ordine al quale si deve arrestare la formula di Mac Laurin nei vari casi.

a) Dalla formula di Mac Laurin abbiamo $\sin x = x - x^3/3! + \cdots$ e quindi

$$\sin x - x = -\frac{1}{6}x^3 + o(x^4) .$$

Dunque l'ordine cercato è $\alpha = 3$.

b) Dalla formula di Mac Laurin abbiamo

$$\sin x = x - \frac{1}{6}x^3 + o(x^4) ,$$

da cui

$$(\sin x)^2 - x^2 = \left[x - \frac{1}{6}x^3 + o(x^4)\right]^2 - x^2 =$$

(è inutile calcolare tutti i termini del quadrato $[...]^2$!)

$$= \left[x^2 - \frac{1}{3}x^4 + o(x^4)\right] - x^2 = -\frac{1}{3}x^4 + o(x^4) .$$

Si conclude che $\alpha = 4$.

c) Usiamo la formula di Mac Laurin per $\log(1-x)$ ed e^{x^2}:

$$\log(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} + o(x^3) ;$$

$$e^{x^2} = 1 + x^2 + \frac{x^4}{2} + o(x^4) .$$

Allora

$$x\log(1-x) + e^{x^2} - 1 =$$

$$= -x^2 - \frac{x^3}{2} - \frac{x^4}{3} + o(x^4) + 1 + x^2 + \frac{x^4}{2} + o(x^4) - 1 =$$

$$= -\frac{x^3}{2} + o(x^3) .$$

Dunque l'ordine cercato è $\alpha = 3$.

d) Usiamo la formula di Mac Laurin per $\cos 2\sqrt{x}$ ed e^{-2x}:

$$\cos 2\sqrt{x} = 1 - 2x + \frac{2}{3}x^2 + o(x^2) ;$$

$$e^{-2x} \quad = 1 - 2x + 2x^2 + o(x^2) .$$

Allora

$$\cos 2\sqrt{x} - e^{-2x} = 1 - 2x + \frac{2}{3}x^2 + o(x^2) - 1 + 2x - 2x^2 + o(x^2) =$$

$$= -\frac{4}{3}x^2 + o(x^2) ,$$

e l'ordine cercato è $\alpha = 2$.

Praticamente la stessa procedura può essere impiegata per il calcolo di numerosi limiti.

***Esempio* 3.6** - Si vogliano calcolare i seguenti limiti:

a) $\lim_{x\to+\infty} x\left[\log\left(1 - \frac{1}{\sqrt{x}}\right) + \frac{1}{\sqrt{x}}\right]$

b) $\lim_{x\to+\infty} x\left[\left(1 + \frac{1}{x}\right)^x - e\right]$.

a) Cambiamo variabile, ponendo $1/x = t$. Allora il limite diventa

$$\lim_{t\to0_+} \frac{1}{t}[\log(1 - \sqrt{t}) + \sqrt{t}] .$$

Usiamo la formula di Mac Laurin per $\log(1 - \sqrt{t})$:

$$\log(1 - \sqrt{t}) = -\sqrt{t} - \frac{t}{2} + o(t) .$$

Dunque:

$$\lim_{t\to0_+} \frac{1}{t}[\log(1 - \sqrt{t}) + \sqrt{t}] = \lim_{t\to0_+} \frac{1}{t}\left[-\sqrt{t} - \frac{t}{2} + o(t) + \sqrt{t}\right] =$$

$$= \lim_{t\to0_+}\left[-\frac{1}{2} + o(1)\right] = -\frac{1}{2} .$$

b) Cambiamo variabile: $1/x = t$.
Il limite diventa:

$$\lim_{t\to0_+} \frac{1}{t}[(1 + t)^{\frac{1}{t}} - e] .$$

Ora scriviamo

$$(1+t)^{\frac{1}{t}} - e = e^{\frac{1}{t}\log(1+t)} - e = e[e^{\frac{1}{t}\log(1+t)-1} - 1] .$$

Usiamo la formula di Mac Laurin per $\log(1+t)$ ad esponente:

$$\log(1+t) = t - \frac{t^2}{2} + o(t^2) .$$

Si ottiene:

$$(1+t)^{\frac{1}{t}} - e = e[e^{\frac{1}{t}[t-\frac{t^2}{2}+o(t^2)]-1} - 1] = e[e^{1-\frac{t}{2}+o(t)-1} - 1] =$$
$$= e[e^{-\frac{t}{2}+o(t)} - 1] .$$

Usiamo ora la formula di Mac Laurin per $e^{-\frac{t}{2}+o(t)}$:

$$e^{-\frac{t}{2}+o(t)} = 1 + \left[-\frac{t}{2} + o(t)\right] + o(t) .$$

Si ottiene:

$$(1+t)^{\frac{1}{t}} - e = e[e^{-\frac{t}{2}+o(t)} - 1] = e\left[1 - \frac{t}{2} + o(t) - 1\right] + o(t) =$$
$$= -\frac{e}{2}t + o(t) .$$

In conclusione:

$$\lim_{t\to 0_+} \frac{1}{t}[(1+t)^{\frac{1}{t}} - e] = \lim_{t\to 0_+} \left[-\frac{e}{2} + o(1)\right] = -\frac{e}{2} .$$

C. Il calcolo numerico di alcune funzioni irrazionali può essere ricondotto al calcolo di un polinomio, mediante l'uso della formula di Taylor.

Un primo problema è il seguente. Sia assegnata una funzione f derivabile $(n+1)$ volte in un intervallo del tipo $(x_0-\delta, x_0+\delta)$. Dato $x \in (x_0-\delta, x_0+\delta)$, si fornisce un valore approssimato di $f(x)$ con la formula

$$f(x) \simeq T_n(x)$$

dove T_n è il polinomio di Taylor con centro x_0, di ordine n, generato da f.

Problema: dare una valutazione dell'errore commesso.

La risposta si basa sulla formula per il resto nella forma di Lagrange:

$$E_n(x) = f(x) - T_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-x_0)^{n+1}$$

dove c è compreso tra x ed x_0.

Occorre levare l'incertezza dovuta alla mancanza di informazione riguardo al punto c. Ciò si può fare se, ad esempio, si conosce una maggiorazione del tipo

$$|f^{(n+1)}(t)| \leq M \tag{3.6}$$

valida per ogni t nell'intervallo chiuso di estremi x, x_0.

In tal caso si avrà:

$$|E_n(x)| \leq \frac{M}{(n+1)!}|x - x_0|^{n+1} . \tag{3.7}$$

***Esempio* 3.7** - Valutiamo l'errore nell'approssimazione di e^{-1} col polinomio di Mac Laurin di 4° grado per la funzione $f(x) = e^x$.

Abbiamo:

$$e^{-1} \simeq 1 - 1 + \frac{1}{2} - \frac{1}{6} + \frac{1}{24} = 0.375 .$$

Poiché $f^{(k)}(x) = e^x$, $\forall k \in \mathbb{N}$, si può scrivere:

$$E_4(-1) = \frac{e^c}{5!}(-1)^5 = -\frac{e^c}{120} \quad \text{con } c \in (-1, 0) .$$

Poiché $0 < e^t \leq 1$ se $t \in [-1, 0]$ si ha:

$$0 > E_4(-1) \geq \frac{-1}{120} = -0,008\overline{3} .$$

In conclusione si ha:

$$0.375 \geq e^{-1} \geq 0.375 - 0.008\overline{3} = 0.366\ldots .$$

Cioè $e^{-1} = 0.3\ldots$ (una sola cifra esatta è garantita).

Un altro problema che si può presentare è il seguente.

Data una funzione derivabile infinite volte in $(x_0 - \delta, x_0 + \delta)$ ed un punto $x \in (x_0 - \delta, x_0 + \delta)$, determinare il polinomio di Taylor T_n di minimo ordine n centrato in x_0 in modo che $T_n(x)$ approssimi $f(x)$ con un errore non superiore ad un valore assegnato α.

Se vale la (3.7) basterà cercare n in modo che

$$\frac{M}{(n+1)!}|x - x_0|^{n+1} \leq \alpha .$$

***Esempio* 3.8** - Considerando, come nell'esempio precedente, $f(x) = e^x$, cerchiamo di approssimare e^{-1} con un errore non superiore ad $\alpha = 10^{-4}$.

Poiché

$$|E_n(-1)| = \frac{e^c}{(n+1)!} \leq \frac{1}{(n+1)!}$$

basterà cercare n in modo che $1/(n+1)! \leq 1/10^4$ ovvero $(n+1)! \geq 10^4$.

Il più piccolo n per cui $(n+1)! \geq 10^4$ è 7; infatti $7! = 5040$, $8! = 40320$.

Dunque il valore

$$T_7(-1) = 1 - 1 + \frac{1}{2} - \frac{1}{6} + \frac{1}{24} - \frac{1}{120} + \frac{1}{720} - \frac{1}{5040} = 0.367857\ldots$$

approssima e^{-1} a meno di 10^{-4}, più esattamente con un errore minore di $1/40320$.

Quindi essendo $E_7(-1) > 0$:

$$0.367857 < e^{-1} < 0.367857 + 0.000024 = 0.36788\ldots$$

da cui si ricava $e^{-1} = 0.3678\ldots$, cioè e^{-1} è noto con 4 cifre decimali esatte.

***Esempio* 3.9** - *Irrazionalità di "e".*

Con l'utilizzo della formula di Mac Laurin generata da e^x si può dimostrare in maniera semplice ed elegante che "e" è un numero irrazionale.

Dalla formula arrestata all'ordine n abbiamo, ricordando che $2 < e < 3$:

$$0 < e - T_n(1) = \frac{e^c}{(n+1)!} < \frac{3}{(n+1)!} \ . \tag{3.8}$$

Moltiplicando la (3.8) per $n!$ otteniamo:

$$0 < en! - n!\left(1 + \frac{1}{1!} + \frac{1}{2!} + \cdots + \frac{1}{n!}\right) < \frac{3}{n+1} \ . \tag{3.9}$$

Sia e razionale, cioè $e = p/q$ con $p, q \in \mathbb{N}_+$; scegliendo nella (3.9) n abbastanza grande in modo che $n > 3$ e in modo che $en!$ sia intero, si avrebbe che l'intero positivo $en! - n!(1 + 1/1! + \cdots + 1/n!)$ sarebbe minore di $3/4$. Assurdo.

Esercizi

10. Determinare i punti stazionari della funzione $f(x) = 3\cos 2x - \sin 3x$ e la loro natura.

11. Dimostrare che se $\varepsilon(x) \to 0$ per $x \to 0$, allora

$$\frac{1}{1 - \varepsilon(x)} = 1 + \varepsilon(x) + \varepsilon(x)^2 + o(\varepsilon(x)^2) \ .$$

12. Trovare l'ordine di infinitesimo per $x \to 0$ delle funzioni seguenti:

$$\text{a) } \frac{1}{1 + (\sin x)^2} - \cos x \quad \text{b) } (\operatorname{arctg} x)^2 - \alpha \log(1 + x^2) \quad (\alpha \in \mathbb{R}) \ .$$

13. Calcolare i seguenti limiti:

a) $\lim_{x\to 0} \frac{1}{x}[(1+\sin x)^{\frac{1}{x}} - e]$ b) $\lim_{x\to +\infty} x^{\alpha}\left(\text{Sh}\,\frac{1}{x} - \sin\frac{1}{x}\right)$ $(\alpha \in \mathbb{R}_+)$

c) $\lim_{x\to 1} \frac{\text{tg}\,\pi x - \pi \log x}{\sin(x-1)^{\alpha}}$ $(\alpha \in \mathbb{R}_+)$ d) $\lim_{x\to +\infty} \frac{x}{(\log x)^2}[x(\sqrt[x]{x} - 1) - \log x]$.

14. Sia f tre volte derivabile in a. Calcolare il limite

$$\lim_{h\to 0} \frac{f(a+h) - f(a-h) - 2hf'(a)}{h^3} .$$

15. Calcolare π con 5 cifre decimali esatte utilizzando la formula di Mac Laurin per l'arcotangente e la seguente identità dovuta a *Gauss*:

$$\pi = 48\,\text{arctg}\left(\frac{1}{8}\right) + 32\,\text{arctg}\left(\frac{1}{57}\right) - 20\,\text{arctg}\left(\frac{1}{239}\right) .$$

16. Se si vuole calcolare sin 1 e cos 1 con h cifre decimali esatte utilizzando la formula di Mac Laurin per $\sin x$ e $\cos x$, qual è l'ordine minimo a cui occorre arrestare la formula?

17.* Dimostrare la proposizione 3.4.
[Suggerimento: utilizzare la formula di Taylor di ordine 2: scrivere

$$f(x) - [f(x_0) - f'(x_0)(x - x_0)] = \frac{f''(x_0)}{2}(x - x_0)^2 + o((x - x_0)^2)$$

e valutare il segno del 1° membro].

3.3 Determinazione del grafico di una funzione

Anche se l'avvento dei mezzi elettronici di calcolo ha reso possibile la costruzione automatica dei grafici di funzioni (anche complicate) riteniamo ugualmente utile che lo studente sappia destreggiarsi manualmente, armato unicamente di carta e matita, in questo classico esercizio.

A tale scopo elenchiamo alcuni tra i punti che possono essere richiesti nello studio del grafico di una funzione assegnata mediante un'espressione analitica $y = f(x)$.

1. Determinazione del dominio di f. Ricerca di eventuali simmetrie e periodicità. Determinazione del segno di f.

2. Comportamento in prossimità dei punti di frontiera di dom(f). Tecnicamente si tratta di calcolare i limiti eventualmente solo destri o sinistri, per x tendente a tali punti. In questa fase si determinano anche eventuali asintoti.

3. Ricerca di eventuali massimi e minimi, locali o globali. Tecnicamente si tratta di determinare il sottoinsieme di dom(f) dove f è derivabile, cioè dom(f'); si cercano in dom(f') i punti stazionari e si determina la loro natura in base allo

studio del segno di f' e cioè della monotonia di f, oppure ricorrendo al teorema 3.5. A volte è possibile riconoscere la natura dei punti stazionari sulla base delle informazioni contenute in 2. Può essere utile calcolare i limiti di f' nei punti di frontiera del suo dominio: si trovano così le pendenze limiti della curva in quei punti e si determinano eventuali cuspidi e punti angolosi, o flessi a tangente verticale.

4. Studio della convessità o concavità di f. Si effettua tramite il calcolo della derivata seconda di f. Una volta individuato $\text{dom}(f'')$, si studia il segno di f''. Si trovano così eventuali punti di flesso (a tangente non verticale).

***Esempio* 3.10** - Mettiamo in pratica lo schema 1-4 per la seguente funzione:

$$f(x) = xe^{\frac{1}{1-|x|}} .$$

1. $\text{dom}(f) = (-\infty,-1) \cup (-1,1) \cup (1,+\infty)$.

La funzione è dispari ($f(x) = -f(-x)$), dunque è sufficiente lo studio del grafico per $x \geq 0$, dove $f(x) = xe^{1/(1-x)} \geq 0$. Notiamo che $f(0) = 0$.

2. Si ha:

$$\lim_{x\to+1_+} xe^{\frac{1}{1-x}} = 0 \ , \quad \lim_{x\to+1_-} xe^{\frac{1}{1-x}} = +\infty \ .$$

L'ultimo limite segnala la presenza di un asintoto verticale (da sinistra) di equazione $x = 1$

$$\lim_{x\to+\infty} xe^{\frac{1}{1-x}} = +\infty \ .$$

Poiché

$$e^{\frac{1}{1-x}} = 1 + \frac{1}{1-x} + o\left(\frac{1}{x}\right) \quad \text{per } x \to +\infty$$

si ha:

$$xe^{\frac{1}{1-x}} = x + \frac{x}{1-x} + o(1) = x - 1 + o(1) \quad \text{per } x \to +\infty \ .$$

Dunque $y = x - 1$ è asintoto obliquo per $x \to +\infty$.

3. La funzione è sicuramente derivabile in tutti i punti di $\text{dom}(f)$, tranne eventualmente $x = 0$ per la presenza di $|x|$.

Per $x > 0, x \neq 1$:

$$f'(x) = e^{\frac{1}{1-x}}\left[1 + x\frac{1}{(1-x)^2}\right] = e^{\frac{1}{1-x}}\frac{1-x+x^2}{(1-x)^2} \ .$$

Poiché $f'(x) > 0 \ \forall x$, si ha che f è strettamente crescente in $(-1,1)$ ed in $(1,+\infty)$ (*non* in $\mathbb{R}_+ \setminus \{1\}$!).

Poiché $\lim_{x\to 0_+} f'(x) = e$, si deduce che, per simmetria, $\lim_{x\to 0_-} f'(x) = e$ per cui $x = 0$ è un punto appartenente a $\mathrm{dom}(f')$ e $f'(0) = e$.

4. Per $x > 0, x \neq 1$, si ha:

$$f''(x) = e^{\frac{1}{1-x}} \frac{2-x}{(1-x)^4} \ .$$

Dunque per $x > 2$ f è concava; per $0 < x < 2$ è convessa; il punto $x = 2$ è punto di flesso; essendo $f(2) = 2/e$ e $f'(2) = 3/e$ l'equazione della retta tangente nel punto di flesso è

$$y = \frac{2}{e} + \frac{3}{e}(x-2) = \frac{3}{e}x - \frac{4}{e} \ .$$

Per simmetria si deduce che anche $x = 0$ è punto di flesso con tangente $y = ex$.

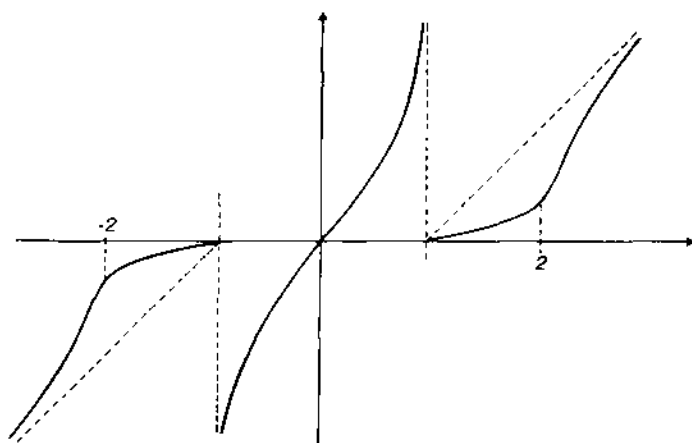



Fig. 6.28 Grafico qualitativo di $y = xe^{\frac{1}{1-|x|}}$.

***Esempio* 3.11** - Si voglia tracciare il grafico di $f(x) = \sqrt{x}\left|1 + \frac{1}{\log x}\right|$.

1. $\mathrm{dom}(f) = (0,1) \cup (1, +\infty)$. Notiamo che $f(x) \geq 0$, e $f(x) = 0$ in $x = e^{-1}$.

2. $\lim_{x\to 0_+} f(x) = 0$; $\lim_{x\to 1_\pm} f(x) = +\infty$; $\lim_{x\to +\infty} f(x) = +\infty$.

C'è un asintoto verticale di equazione $x = 1$, non vi sono asintoti obliqui poiché $f(x) \sim \sqrt{x}$ per $x \to +\infty$.

3. L'unico punto di non derivabilità nel dominio di f può essere solo $x = 1/e$ dove si annulla l'argomento del modulo.

Poiché

$$\frac{d}{dx}\left(1+\frac{1}{\log x}\right) = -\frac{1}{x(\log x)^2}$$

che in $1/e$ vale $-e$ si ricava che $(1/e, 0)$ è effettivamente un punto angoloso.

Si ha

$$f'_+\left(\frac{1}{e}\right) = e \ , \ f'_-\left(\frac{1}{e}\right) = -e \ .$$

Negli altri punti f è derivabile e

$$f'(x) = \text{sign}\left(1+\frac{1}{\log x}\right)\frac{(\log x)^2+\log x-2}{2\sqrt{x}(\log x)^2} \ .$$

Dunque $f'(x) = 0$ se $(\log x)^2 + \log x - 2 = 0$ e cioè se $x = e^{-2}$ o $x = e$.

È facile verificare (dai punti 1. e 2.) che $x = e^{-2}$ è punto di massimo relativo, con $f(e^{-2}) = 1/2e$ mentre $x = e$ è di minimo relativo con $f(x) = 2\sqrt{e}$. Notiamo inoltre che $f'(x) \to +\infty$ se $x \to 0_+$.

Un grafico qualitativo di f è il seguente (*).

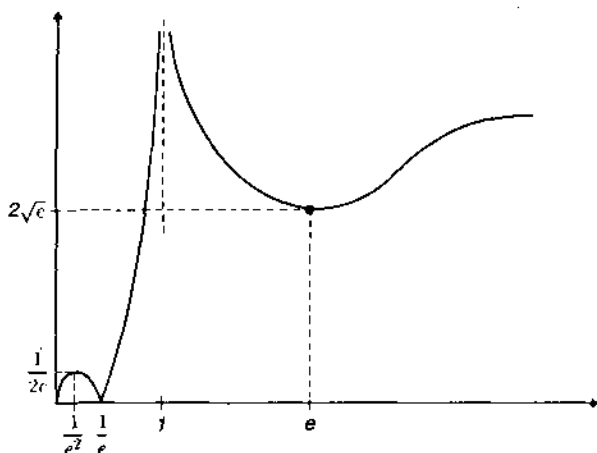



Fig. 6.29 Grafico qualitativo di $y = \sqrt{x}\left|1+\frac{1}{\log x}\right|$.

3.4 Risoluzione numerica di equazioni

Presentiamo qui alcuni metodi per risolvere numericamente un'equazione del tipo

(*) Alternativamente si sarebbe potuto procedere studiando la funzione $g(x) = \sqrt{x}\left(1+\frac{1}{\log x}\right)$ e osservare che $f = |g|$.

$$f(x) = 0 \,. \tag{3.10}$$

Questi metodi si rivelano necessari qualora non si conoscano formule esplicite per le radici di (3.10) oppure qualora le formule stesse siano troppo complicate.

Si troveranno in generale solo approssimazioni delle soluzioni esatte ed uno dei problemi principali è quello di avere informazioni sull'errore di approssimazione.

Dal punto di vista pratico occorre sottolineare che la quasi totalità dei metodi usati prevede la ripetizione di cicli di calcolo (*metodi iterativi*) che ben si adattano all'uso di un elaboratore; d'altro canto proprio l'uso di un mezzo automatico introduce inevitabilmente ad ogni passo un errore dovuto agli arrotondamenti (o nel peggiore dei casi, ai troncamenti) effettuati dalla macchina.

Si presenta così il problema di cercare *algoritmi* di calcolo che evitino di propagare o esaltare tali errori ovvero di costruire algoritmi *stabili*.

Pur non desiderando entrare in dettaglio in questo importante problema dell'Analisi numerica, segnaliamo al lettore che anche algoritmi estremamente semplici possono essere instabili; ad esempio il ciclo di calcolo rappresentato dalla seguente successione per ricorrenza

$$s_0 = \frac{1}{77} \ , \ s_{n+1} = 78 s_n - 1$$

è instabile. Per convincersene basta osservare che $s_n = 1/77$ per ogni n, ma, usando una calcolatrice, dopo poche iterazioni si troverebbero risultati assurdi.

Il primo metodo che esaminiamo è dovuto a Newton e si chiama anche *metodo delle tangenti*.

Consideriamo l'equazione $f(x) = 0$ e supponiamo di aver accertato l'esistenza e l'unicità di una soluzione r all'interno di un intervallo $[a, b]$. Si dice che si è *isolata* la radice r.

A questo proposito si potrà ricorrere ad esempio ad uno studio qualitativo del grafico di f.

Illustriamo il metodo di Newton in riferimento alla figura 6.30.

L'idea è di costruire una successione $\{x_n\}$ che sia convergente ad r.

Ogni x_n può essere considerato come un'approssimazione di r stessa.

A questo scopo si parte da $x_0 = a$ e "*si linearizza l'equazione* $f(x) = 0$ *in quel punto*"; si sostituisce cioè ad $y = f(x)$ l'equazione della retta tangente al grafico di f nel punto $(x_0, f(x_0))$, ovvero $y = f(x_0) + f'(x_0)(x - x_0)$.

Invece di risolvere $f(x) = 0$ si risolve $f(x_0) + f'(x_0)(x - x_0) = 0$, che ha per soluzione

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} \,.$$

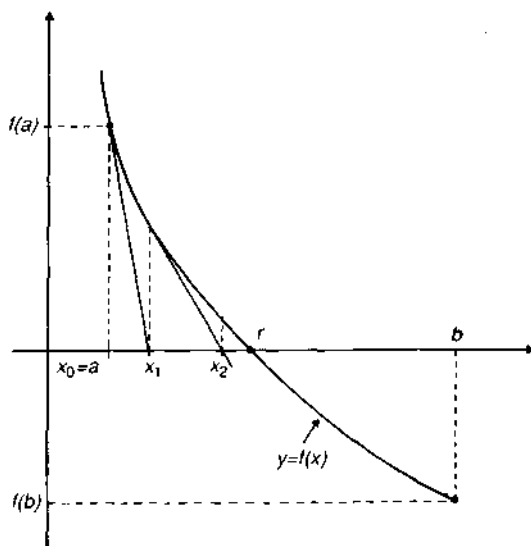



Fig. 6.30

Il procedimento si ripete con x_1 al posto di x_0, linearizzando l'equazione $f(x) = 0$ in $x = x_1$, sostituendola cioè con l'equazione

$$f(x_1) + f'(x_1)(x - x_1) = 0$$

che ha per soluzione

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} .$$

Iterando n volte il procedimento si perviene alla seguente successione per ricorrenza

$$x_0 = a \quad x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})} . \tag{3.11}$$

Nella situazione descritta in figura 6.30, è intuitivo che $x_n \to r_-$ per $n \to +\infty$; qui f è *convessa* e f' *non si annulla mai.*

D'altra parte, se f non è convessa o concava in $[a, b]$ il procedimento può arrestarsi come nel caso illustrato in figura 6.31. Partendo da $x_0 = a$ in figura 6.31 il procedimento si arresta al secondo passo.

Una formula identica alla (3.11) si otterrebbe partendo da $x_0 = b$ anziché da a. Per evitare inconvenienti è opportuno, nel caso di funzioni convesse o concave, derivabili 2 volte, partire da un punto iniziale x_0 tale che $f(x_0)$ abbia lo stesso segno di $f''(x_0)$.

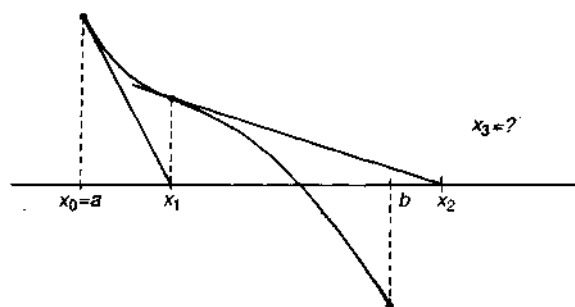



Fig. 6.31

Formalizziamo la precedente discussione nel seguente teorema.

▪ **Teorema 3.6** - *Sia $f \in C^2([a,b])$ tale che:*

i) esiste un solo $r \in (a,b)$ tale che $f(r) = 0$.

ii) $\forall x \in [a,b]$, $f'(x) \neq 0$ e $f''(x) > 0$ oppure $f''(x) < 0$.

Allora la seguente successione

$$x_0 = a \quad se \quad f(a)f''(a) > 0$$

oppure

$$x_0 = b \quad se \quad f(b)f''(b) > 0$$

e

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})}$$

converge a r per $n \to +\infty$.

Dimostrazione - Esaminiamo solo il caso $f(a)$ ed $f''(a)$ positive; gli altri casi sono analoghi. Abbiamo dunque che $f''(x) > 0$ e $f'(x) < 0$ $\forall x \in [a,b]$. Ne segue che, per ogni $n \geq 1$, $x_{n-1} < x_n < r$.

La successione $\{x_n\}$ è dunque monotona crescente e limitata. Perciò è convergente ad un limite $l \leq r$.

Passando al limite nella relazione di ricorrenza (3.11) si ottiene allora, poiché sia f, sia f' sono continue:

$$l = l - \frac{f(l)}{f'(l)}$$

e cioè $f(l) = 0$. Dalla *i*) si deduce $l = r$. □

***Esempio* 3.12** - Consideriamo l'equazione

$$f(x) = e^{-x} - x = 0 \ .$$

Confrontando i grafici di $y = e^{-x}$ e $y = x$ si ottiene che esiste un'unica radice r compresa tra 0 ed 1.

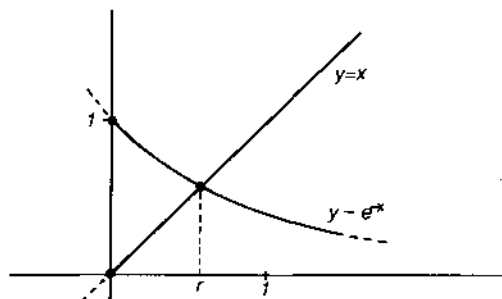



Fig. 6.32

Abbiamo

$$f'(x) = -e^{-x} - 1 < 0 \quad \forall x \in [0,1]$$
$$f''(x) = e^{-x} > 0 \qquad \forall x \in [0,1] \,.$$

Inoltre $f(0) = 1 > 0$ e $f(1) = 1/e - 1 < 0$.

Per il calcolo di r poniamo dunque:

$$x_0 = 0 \quad x_n = x_{n-1} + \frac{e^{-x_{n-1}} - x_{n-1}}{e^{-x_{n-1}} + 1} = \frac{x_{n-1} + 1}{e^{x_{n-1}} + 1} \,.$$

Usando una comune calcolatrice abbiamo trovato, considerando 6 cifre decimali dopo il punto:

$$x_1 = 0.5$$
$$x_2 = 0.566311$$
$$x_3 = 0.567143$$
$$x_4 = 0.567143$$

Da x_3 in poi le 6 cifre si stabilizzano. Notiamo che $e^{-x_4} - x_4 \simeq 4.551137 \cdot 10^{-7}$. Un risultato buono dopo solo 3 iterazioni.

Studiamo un po' da vicino la rapidità di convergenza introducendo l'errore $E_n = |r - x_n|$ e confrontandolo con quello al passo successivo E_{n+1}.

Abbiamo, essendo $f(r) = 0$:

$$\begin{aligned} x_{n+1} - r &= x_n - r - \frac{f(x_n) - f(r)}{f'(x_n)} = \\ &= \frac{(x_n - r)f'(x_n) - f(x_n) + f(r)}{f'(x_n)} \,. \end{aligned} \tag{3.12}$$

Dalla formula di Taylor con centro in x_n abbiamo che:

$$f(r) = f(x_n) + f'(x_n)(r - x_n) + \frac{1}{2}f''(c)(r - x_n)^2 \; .$$

dove c è un opportuno punto tra x_n ed r.

Sostituendo nella (3.12) ricaviamo, dopo semplificazioni elementari:

$$x_{n+1} - r = \frac{1}{2}\frac{f''(c)}{f'(x_n)}(x_n - r)^2 \; .$$

Se ora poniamo

$$M = \frac{1}{2}\frac{\max|f''|}{\min|f'|} \; ,$$

otteniamo

$$E_{n+1} = |x_{n+1} - r| \le M(x_n - r)^2 = ME_n^2 \; . \tag{3.13}$$

La (3.13) indica che se E_n è "piccolo", E_{n+1} è "molto più piccolo".

Iterando la (3.13) otteniamo:

$$\begin{aligned} E_{n+1} &\le M \cdot E_n^2 \le M \cdot M^2 \cdot E_{n-1}^4 \le M \cdot M^2 \cdot M^4 \cdot E_{n-2}^8 \le \dots \le \\ &\le M^{1+2+4+\dots+2^{n-1}} \cdot E_0^{2^n} \; . \end{aligned}$$

ovvero

$$E_{n+1} \le M^{2^n - 1} E_0^{2^n} \; . \tag{3.14}$$

La (3.14), scritta nella forma $|x_{n+1} - r| \le M^{-1}(M|x_0 - r|)^{2^n}$, indica che se $M|x_0 - r| < 1$, cioè se siamo partiti "abbastanza vicino" ad r, l'errore decresce a zero con velocità più che esponenziale.

Una variante del metodo di Newton è quello di *falsa posizione* che qui illustriamo senza entrare in troppi dettagli.

L'idea è di sostituire $f'(x_{n-1})$ nella formula (3.11) con il rapporto incrementale

$$\frac{f(x_{n-1}) - f(x_{n-2})}{x_{n-1} - x_{n-2}} \; .$$

Naturalmente per far partire l'iterazione occorrono 2 punti iniziali x_0, x_1; la formula iterativa è la seguente:

$$x_n = x_{n-1} - \frac{x_{n-1} - x_{n-2}}{f(x_{n-1}) - f(x_{n-2})} f(x_{n-1}) \; . \tag{3.15}$$

Geometricamente significa *linearizzare* l'equazione $f(x) = 0$ con una retta secante anziché una tangente come illustrato in figura 6.33.

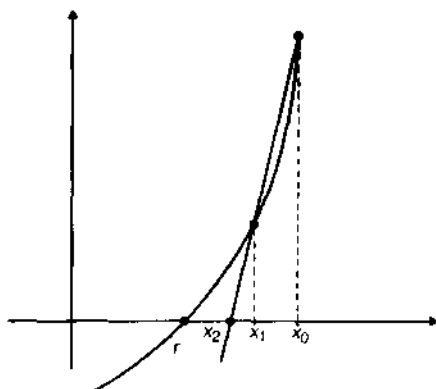



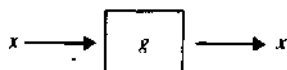
Fig. 6.33

Vediamo ora il metodo delle *approssimazioni successive* la cui importanza va ben oltre il calcolo numerico delle radici di un'equazione e che verrà inquadrato nel secondo volume nel suo ambiente naturale, quello degli spazi metrici.

In riferimento all'equazione $f(x) = 0$, aggiungendo x ad entrambi i membri e ponendo $g(x) = f(x) + x$, ci si riporta a risolvere l'equazione

$$g(x) = x \; . \tag{3.16}$$

Geometricamente risolvere la (3.16) equivale a determinare l'intersezione tra i grafici di $y = x$ e $y = g(x)$.

Analiticamente, le soluzioni della (3.16) *sono numeri che vengono trasformati in sé stessi da* g. Tali numeri si chiamano **punti fissi** di g:

Il metodo delle *approssimazioni successive* consiste nella costruzione di una successione $\{x_n\}$ secondo lo schema iterativo seguente:

si fissa un valore (opportuno) iniziale x_0; si pone poi:

$$x_{n+1} = g(x_n) \quad n \geq 0 \; . \tag{3.17}$$

Se ... "tutto va bene", se cioè si riesce a dimostrare che x_n converge ad un numero reale l e se g è continua, passando al limite nella (3.17) si trova

$$l = g(l)$$

e cioè che l è un punto fisso di g.

L'andamento della successione $\{x_n\}$ può essere efficacemente visualizzato seguendo il percorso dei punti P_n di coordinate (x_n, x_{n+1}) come mostrato in figura 6.34.

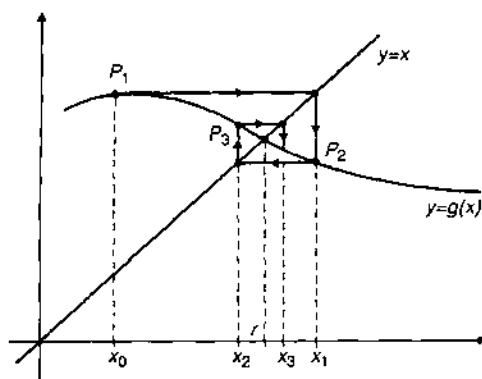



Fig. 6.34

In figura i punti P_n convergono al punto "fisso" P di coordinate (r, r) con un tipico percorso a spirale.

Il prossimo teorema indica che la convergenza della successione (3.17) dipende dalla pendenza di g.

■ **Teorema 3.7** - *Sia $g: [a,b] \to [a,b]$, derivabile in $[a,b]$ e tale che esista un numero reale $k, 0 < k < 1$, tale che*

$$|g'(x)| \leq k \quad \forall x \in [a,b] \,. \tag{3.18}$$

Allora g ha uno ed un solo punto fisso r in $[a,b]$ e la successione (3.17) *converge ad r indipendentemente dal valore iniziale scelto $x_0 \in [a,b]$.*

Dimostrazione - Facciamo vedere che la successione $\{x_n\}$ soddisfa la condizione di Cauchy e pertanto risulta convergente ad un limite finito r che per quanto abbiamo già osservato è un punto fisso di g.

Applicando il teorema di Lagrange a g nell'intervallo di estremi x_n ed x_{n+1} abbiamo, per la (3.18)

$$|x_{n+1} - x_n| = |g(x_n) - g(x_{n-1})| = |g'(c)| \cdot |x_n - x_{n-1}| \leq k|x_n - x_{n-1}|, \tag{3.19}$$

dove c è un opportuno punto tra x_n ed x_{n+1}, in generale dipendente da n .

Iterando la (3.19) si ottiene

$$|x_{n+1} - x_n| \leq k^2|x_{n+1} - x_{n-2}| \leq \ldots \leq k^n|x_1 - x_0| \,. \tag{3.20}$$

Siano $p, n \in \mathbb{N}$. Dalla disuguaglianza triangolare e dalla (3.20) si ha:

$$\begin{aligned}|x_{n+p} - x_n| &\leq |x_{n+p} - x_{n+p-1}| + |x_{n+p-1} - x_{n+p-2}| + \cdots + |x_{n+1} - x_n| \leq \\ &\leq k^{n+p-1}|x_1 - x_0| + k^{n+p-2}|x_1 - x_0| + \cdots + k^n|x_1 - x_0| = \\ &= k^n(k^{p-1} + k^{p-2} + \cdots + 1)|x_1 - x_0| = \frac{1-k^p}{1-k}k^n|x_1 - x_0| < \\ &< \frac{k^n}{1-k}|x_1 - x_0| \,,\end{aligned}$$

ricordando che

$$\sum_{s=0}^{p-1} k^s = \frac{1-k^p}{1-k} .$$

Essendo $k < 1$ si ha $k^n \to 0$ per $n \to +\infty$ e quindi la successione $\{x_n\}$ soddisfa la condizione di Cauchy.

Per dimostrare che il punto fisso r è unico, supponiamo per assurdo che r_1 sia un altro punto fisso: $g(r_1) = r_1$. Allora, sempre per la (3.18),

$$|r_1 - r| = |g(r_1) - g(r)| \leq k|r_1 - r| < |r_1 - r| .$$

che è una contraddizione. □

Osservazione 3.2 - Nel corso della dimostrazione abbiamo visto che, siccome $|g'(x)| \leq k < 1 \ \forall x \in [a,b]$, per ogni coppia di punti $x_1, x_2 \in [a,b]$ si ha:

$$|g(x_1) - g(x_2)| \leq k|x_1 - x_2| . \tag{3.21}$$

Essendo $k < 1$ possiamo interpretare la condizione (3.21) come *effetto di contrazione* delle distanze operato dalla funzione g. In altri termini la (3.21) dice che "*la distanza tra le immagini (tramite g) di x_1 ed x_2 è contratta di un fattore k rispetto alla distanza tra x_1 ed x_2*".

Una funzione g che soddisfi la (3.21) per ogni coppia di punti del suo dominio si dice *contrazione di costante k*.

Dunque se vale la (3.18) g è una contrazione di costante k.

Osservazione 3.3 - Nella dimostrazione del teorema abbiamo trovato la disuguaglianza

$$|x_{n+p} - x_n| < \frac{k^n}{1-k}|x_1 - x_0| , \tag{3.22}$$

valida per ogni $n, p \geq 0$.

Facciamo tendere p a $+\infty$ nella (3.22); essendo $x_{n+p} \to r$ si ottiene

$$|x_n - r| < \frac{k^n}{1-k}|x_1 - x_0| .$$

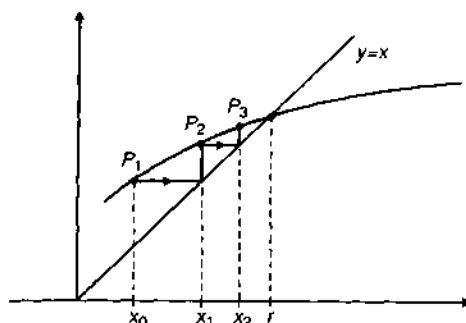



Fig. 6.35

Tale disuguaglianza ci dà informazioni sulla velocità con la quale x_n converge ad r: più piccolo è k (la pendenza di g) e più veloce sarà la convergenza.

Si può altresì osservare che se $-k \leq g'(x) \leq 0$, $\forall x \in [a,b]$ la convergenza avviene come in figura 6.34, con un percorso a spirale: i punti x_n sono alternativamente maggiori o minori di r.

Se invece $0 \leq g'(x) \leq k$, $\forall x \in [a,b]$ la convergenza avviene in maniera monotona come in figura 6.35.

***Esempio* 3.13** - Consideriamo ancora l'equazione

$$e^{-x} = x \ .$$

Qui $g(x) = e^{-x}$; $g'(x) = -e^{-x}$. Se stiamo discosti da zero, ad esempio in $[1/3, 1]$ si ha $|g'(x)| \leq e^{-1/3} < 1$. Inoltre se $x \in [1/3, 1]$, si ha $1/3 \leq e^{-x} \leq 1$. Dunque

$$g\colon \left[\frac{1}{3}, 1\right] \to \left[\frac{1}{3}, 1\right] .$$

Il metodo delle approssimazioni successive conduce alla successione

$$x_0 = 1 \quad x_{n+1} = e^{-x_n}$$

che dà, sempre accontentandoci di 6 cifre decimali dopo il punto:

$$\begin{array}{lll}
x_1 = 0.367879 & x_6 \ = 0.579612 & \ldots \quad x_{24} = 0.567143 \\
x_2 = 0.692200 & x_7 \ = 0.560115 & \\
x_3 = 0.500473 & x_8 \ = 0.571143 & \\
x_4 = 0.606243 & x_9 \ = 0.564879 & \\
x_5 = 0.545395 & x_{10} = 0.568428 \ . &
\end{array}$$

Dalla 24^a iterazione in poi le prime 6 cifre decimali si stabilizzano. Si vede che per avere la stessa precisione del metodo di Newton occorrono 24 cicli di calcolo anziché ... 3.

Esercizi

18. Verificare che, nel metodo di Newton, $|x_n - r|$ è asintotico a $|x_{n+1} - x_n|$ per $n \to +\infty$.
(Cioè: per n grande è attendibile valutare l'errore di approssimazione al passo n-esimo (ignoto!) con la differenza $|x_{n+1} - x_n|$, che è nota).

19. Verificare la convergenza del metodo di *falsa posizione*, nelle stesse ipotesi del teorema 3.6.

20. Un altro metodo di calcolo approssimato delle radici di un'equazione della forma $f(x) = 0$ è quello **delle corde**.
Le ipotesi richieste ad f sono:

i) $f \in C([a, b])$;

ii) $f(a)f(b) < 0$.

Inoltre esista in $[a, b]$ un'unica radice r.
Considerata la seguente figura:

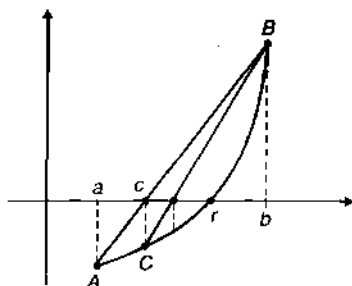



Fig. 6.36

si traccia la retta AB; detta c la sua intersezione con l'asse delle ascisse si ripete il procedimento con la retta CB e così via.
Determinare la formula generale per la successione approssimante e studiarne la convergenza.

21. Scrivere un programma che fornisca una tabella di confronto tra i vari metodi risolutivi visti, relativamente ad un'equazione di cui si conosca la soluzione esatta.

22. Verificare che il metodo di Newton rientra nello schema delle approssimazioni successive per un'opportuna g.

23. Supponiamo che g abbia un'unico punto fisso r e che la successione $x_n = g(x_{n-1})$, x_0 assegnato, converga ad r. Si dice che il metodo è di ordine $\alpha \in \mathbb{R}_+$ se esiste $M > 0$ tale che, $\forall n \in \mathbb{N}$

$$|x_{n+1} - r| \leq M|x_n - r|^\alpha .$$

Dimostrare che, se $g'(r) \neq 0$, il metodo è del 1° ordine mentre se $g'(r) = 0$ $(g''(r) \neq 0)$ il metodo è del 2° ordine.
Verificare che nel caso del metodo di Newton (esercizio 22) si ha $g'(r) = 0$.

24. Risolvere le seguenti equazioni usando il metodo che si ritiene più conveniente:

a) $2\log x + x - 4 = 0$

b) $x^{100} + x - 12 = 0$

c) $2x + 1 - (x - 1)\log x = 0$

d) $e^{x} - x - 2 = 0$

e) $x^{3} = e^{-x} + 1$.

Calcolo differenziale 2
Funzioni di più variabili 7

In questo capitolo estendiamo il calcolo differenziale a funzioni reali di n variabili reali (sez. 1) e a funzioni vettoriali di n variabili reali (sez. 2).

Sia $\mathbf{f}: \mathbb{R}^n \supseteq A \to \mathbb{R}^m$; l'obiettivo è quello di studiare le variazioni di $\mathbf{f}$ relativamente a variazioni del suo argomento nell'intorno di un punto interno ad A.

Per questa ragione, A sarà, salvo avviso contrario, quasi sempre un aperto di $\mathbb{R}^n$.

Nel par. 2.4 è presentato l'importante teorema di inversione locale, generalizzazione al caso di trasformazioni non lineari del teorema di Cramer.

La sezione 3 è dedicata alle funzioni implicite. Il principale risultato (teor. 3.8) è l'analogo non lineare del teorema di Rouché–Capelli.

Il teorema di inversione locale ed il teorema 3.8 giocano un ruolo essenziale in geometria differenziale e nella teoria delle equazioni differenziali, argomenti che saranno trattati nel secondo volume.

1. FUNZIONI DA $\mathbb{R}^n$ IN $\mathbb{R}$

1.1 Derivate direzionali e derivate parziali

Siano $f: \mathbb{R}^n \supseteq A \to \mathbb{R}$, A aperto di $\mathbb{R}^n$, $\mathbf{x} \in A$.

Se $n = 1$, la derivata prima di f assegna il tasso di incremento di f relativamente ad x. Essendoci una sola dimensione, è unica la direzione lungo la quale x viene incrementato. In dimensione $n > 1$ il modo in cui f varia dipende dalla direzione nella quale avviene la variazione di $\mathbf{x}$.

Ciò conduce in modo naturale al concetto di derivata in una data direzione o derivata direzionale, che ora descriviamo.

Introduciamo un versore $\mathbf{v} \in \mathbb{R}^n$ ($\|\mathbf{v}\| = 1$) che assegna una direzione e un verso in $\mathbb{R}^n$ e, per ogni $t \in \mathbb{R}$ tale che $\mathbf{x} + t\mathbf{v} \in A$, consideriamo il rapporto

$$\frac{f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x})}{t}, \tag{1.1}$$

detto *rapporto incrementale di f nella direzione* $\mathbf{v}$, che rappresenta il tasso di variazione medio di f in quella direzione e verso.

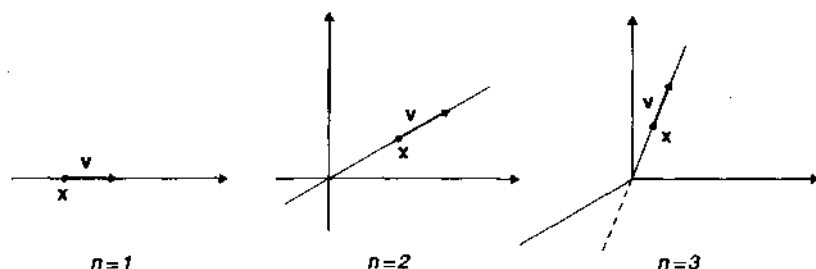



Fig. 7.1

Definizione 1.1 - *Quando esiste finito, il*

$$\lim_{t\to 0} \frac{f(\mathbf{x}+t\mathbf{v})-f(\mathbf{x})}{t}$$

si chiama derivata nella direzione $\mathbf{v}$ *di* f *nel punto* x *e si indica con* $D_{\mathbf{v}}f(\mathbf{x})$.
In questo caso f si dice *derivabile nella direzione* $\mathbf{v}$ in $\mathbf{x}$.

La definizione 1.1 è essenzialmente uni–dimensionale; ciò si può mettere in luce osservando che, essendo $\mathbf{x}$ e $\mathbf{v}$ fissati, posto $\varphi(t) = f(\mathbf{x}+t\mathbf{v})$, φ è una funzione della variabile reale t, definita in un intorno di $t = 0$, e si ha:

$$D_{\mathbf{v}}f(\mathbf{x}) = \varphi'(0)\ .$$

***Esempio* 1.1** - Si voglia calcolare la derivata lungo una direzione $\mathbf{v} \in \mathbb{R}^3$ della funzione $f : \mathbf{x} \mapsto ||\mathbf{x}||^2 : \mathbb{R}^3 \to \mathbb{R}$, nel punto generico $\mathbf{x}$.
Posto $\varphi(t) = ||\mathbf{x}+t\mathbf{v}||^2$ si ha:

$$\varphi'(t) = \frac{d}{dt}\sum_{j=1}^{3}(x_j+tv_j)^2 = 2\sum_{j=1}^{3}(x_j+tv_j)v_j = 2 < \mathbf{x}+t\mathbf{v},\mathbf{v} > \ .$$

Dunque $D_{\mathbf{v}}f(\mathbf{x}) = \varphi'(0) = 2 < \mathbf{x},\mathbf{v} >$ e f è derivabile lungo qualunque direzione in qualunque punto.

Nella definizione di $D_{\mathbf{v}}f(\mathbf{x})$ intervengono solo i valori di f lungo la retta $\mathbf{x}+t\mathbf{v}$ (ovvero la restrizione di f lungo tale retta), cosicché l'esistenza della derivata in una direzione non dà informazioni circa l'esistenza della derivata in un'altra direzione.
Per esempio la funzione

$$f(x_1,x_2) = \frac{x_1^2}{x_1^2+x_2^2} \quad \text{se} \quad (x_1,x_2) \neq (0,0)$$

e $f(0,0) = 0$,

ha $D_{\mathbf{e}^2} f(0,0) = 0$ mentre $D_{\mathbf{e}^1} f(0,0)$ non esiste. (L'allievo è invitato a verificare i calcoli).

Particolare importanza rivestono le derivate lungo le direzioni degli assi coordinati, individuate dai versori $\mathbf{e}^1, \mathbf{e}^2, \dots, \mathbf{e}^n$.

In questo caso $D_{\mathbf{e}^j} f(\mathbf{x})$ prende il nome di **derivata parziale rispetto ad x_j**, di f nel punto $\mathbf{x}$.

Per le derivate parziali si usano le seguenti notazioni:

$$\frac{\partial f}{\partial x_j} \ , \ D_j f \ , \ D_{x_j} f \ , \ \partial_{x_j} f \ , \ f_{x_j} \ .$$

In una derivata parziale, quella rispetto a x_j, per esempio, viene incrementata la sola variabile x_j; ne segue che per il calcolo di f_{x_j} si può pensare alle altre variabili come costanti ed utilizzare le regole di derivazione note per le funzioni di una variabile.

Ad esempio, se

$$f(x_1, x_2, x_3) = x_1 \sqrt{x_2^2 + x_3^2}$$

si ha:

$$f_{x_1}(x_1, x_2, x_3) = \sqrt{x_2^2 + x_3^2}$$

$$f_{x_2}(x_1, x_2, x_3) = \frac{x_1 x_2}{\sqrt{x_2^2 + x_3^2}}$$

$$f_{x_3}(x_1, x_2, x_3) = \frac{x_1 x_3}{\sqrt{x_2^2 + x_3^2}} \ .$$

Se una funzione $f : \mathbb{R}^n \supseteq A \to \mathbb{R}$ ammette n derivate parziali in un punto $\mathbf{x} \in A$, è definito un vettore che si chiama **gradiente** *di f in $\mathbf{x}$, le cui componenti sono le n derivate parziali di f in $\mathbf{x}$ e che si indica con i simboli $\nabla f(\mathbf{x})$ o $\operatorname{grad} f(\mathbf{x})$:*

$$\nabla f(\mathbf{x}) := (f_{x_1}(\mathbf{x}), f_{x_2}(\mathbf{x}), \dots, f_{x_n}(\mathbf{x})) \ .$$

> Il significato geometrico di derivata direzionale nel caso $n = 2$ è illustrato in figura 7.2 dove il piano π è perpendicolare al piano x_1, x_2 e lo interseca nella retta $\mathbf{x} + t\mathbf{v}$. Si ha che $D_{\mathbf{v}} f(\mathbf{x}) = \operatorname{tg} \theta =$ pendenza della sezione del piano π col grafico di f, nel punto di coordinate $(\mathbf{x}, f(\mathbf{x}))$.

Si possono definire derivate destre o sinistre lungo una data direzione individuata da un versore $\mathbf{v}$ in modo ovvio:

$$D^{\pm}_{\mathbf{v}} f(\mathbf{x}) := \lim_{t \to 0\pm} \frac{f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x})}{t}$$

quando i limiti esistano finiti.

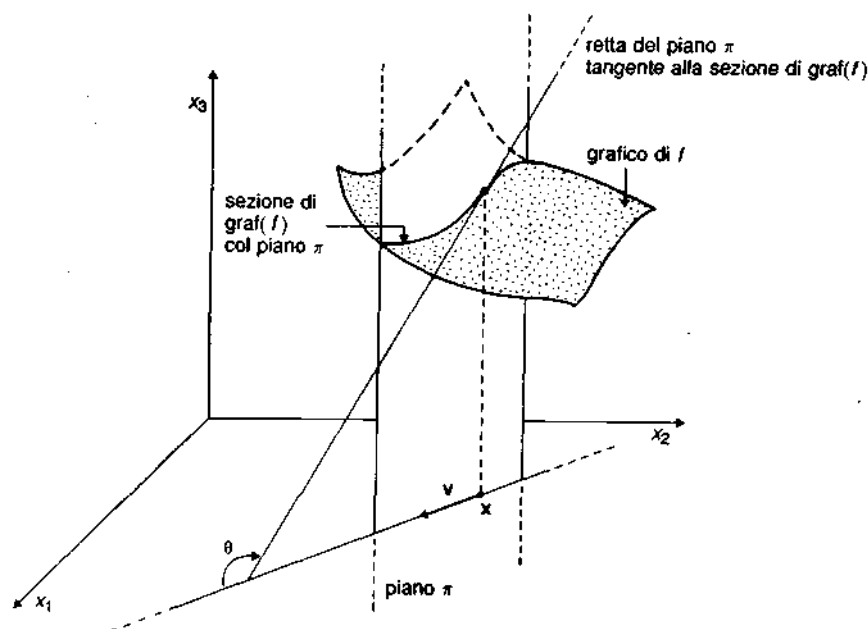



Fig. 7.2

In molte questioni di ottimizzazione si ha a che fare con funzioni $f : \overline{A} \to \mathbb{R}$, dove $\overline{A}$ è la chiusura di un aperto $A \subseteq \mathbb{R}^n$, ed occorre calcolare derivate direzionali in punti $\mathbf{x} \in \partial A$.

Queste derivate sono necessariamente solo destre o sinistre, inoltre è evidente che, diversamente dal caso dei punti interni, avremo solo determinate direzioni (o versori) ammissibili, dipendenti dal punto $\mathbf{x}$.

Infatti per calcolare $D_{\mathbf{v}}^{+} f(\mathbf{x})$ (risp. $D_{\mathbf{v}}^{-} f(\mathbf{x})$) per $\mathbf{x} \in \partial A$, occorre che, per t in un intorno destro (risp. sinistro) di zero, si abbia $\mathbf{x} + t\mathbf{v} \in A$. Nei casi più comuni i versori $\mathbf{v} \in \mathbb{R}^n$ ammissibili costituiscono nel loro insieme un "*cono*" con vertice in $\mathbf{x}$.

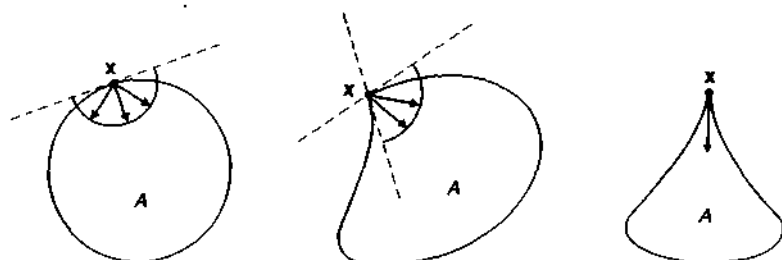



Fig. 7.3 Esempi di direzioni ammissibili in vari casi.

Concludiamo il paragrafo illustrando la relazione tra derivabilità e continuità.
Evidentemente se $D_{\mathbf{v}} f(\mathbf{x})$ esiste allora f è continua lungo la direzione $\mathbf{v}$ in $\mathbf{x}$.
Il seguente esempio mostra che l'esistenza di *tutte* le derivate direzionali in un punto *non implica* la continuità in quel punto, in dimensione maggiore di 1.

***Esempio* 1.2** - Sia $f(x_1, x_2) = \begin{cases} x_1 e^{x_1/x_2} & \text{se } x_2 \neq 0 \\ 0 & \text{se } x_2 = 0 \,. \end{cases}$

Poiché siamo in $\mathbb{R}^2$ una qualunque direzione è individuata dal versore $\mathbf{v}_\theta$ di componenti $(\cos\theta, \sin\theta)$ con $\theta \in [0, 2\pi)$.
Proviamo che f è derivabile lungo qualunque direzione in $(x_1, x_2) = (0,0)$. Si ha:

$$D_{\mathbf{v}_\theta} f(0,0) = \lim_{t \to 0} \frac{f(t\cos\theta, t\sin\theta)}{t} = \lim_{t \to 0} \frac{t\cos\theta e^{\operatorname{cotg}\theta}}{t} =$$

$$= \cos\theta e^{\operatorname{cotg}\theta} \quad (\theta \neq 0)$$

$$D_{\mathbf{v}_0} f(0,0) = f_{x_1}(0,0) = 0 \,.$$

D'altra parte f *non* è continua in $(0,0)$ poiché il limite per $(x_1, x_2) \to (0,0)$ ad esempio lungo la cubica di equazione $x_2 = x_1^3$ è ∞ (infatti $\lim_{x_1 \to 0^\pm} x_1 e^{1/x_1^2} = = \pm\infty$).

L'esempio 1.2 indica che, se $n > 1$, per studiare proprietà relative ad un intorno n–dimensionale di un punto occorre un concetto più potente di quello di derivabilità.
Tale concetto è quello di differenziabilità, oggetto del prossimo paragrafo, che, dunque, in dimensione maggiore di 1, non risulta più equivalente a quello di derivabilità.

1.2 Differenziale

L'idea di differenziabilità è essenzialmente quella di poter approssimare l'incremento $\Delta f = f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x})$ con una funzione (reale) *lineare* in $\mathbf{h}$ a meno di infinitesimi di ordine superiore a $\|\mathbf{h}\|$.
Ricordiamo che ogni funzione lineare $l : \mathbb{R}^n \to \mathbb{R}$ è "rappresentata" da un unico vettore $\mathbf{a} \in \mathbb{R}^n$, nel senso che

$$l(\mathbf{h}) = \langle \mathbf{a}, \mathbf{h} \rangle \qquad \forall\, \mathbf{h} \in \mathbb{R}^n \quad (^*).$$

Siamo così condotti alla seguente definizione.

(*) $\mathbb{R}^n$ si intende riferito alla base canonica.

Definizione 1.2 - *Sia $f : \mathbb{R}^n \supseteq A \to \mathbb{R}$, A aperto; f si dice differenziabile in $\mathbf{x} \in A$ se esiste un vettore $\mathbf{a} \in \mathbb{R}^n$ tale che*

$$f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) = < \mathbf{a},\mathbf{h} > + o(\|\mathbf{h}\|) \quad per \ \|\mathbf{h}\| \to 0 \tag{1.2}$$

per ogni $\mathbf{h} \in \mathbb{R}^n$ con $\mathbf{x}+\mathbf{h} \in A$.

L'applicazione lineare da $\mathbb{R}^n$ in $\mathbb{R}$ data da

$$\mathbf{h} \mapsto < \mathbf{a},\mathbf{h} >$$

prende il nome di **differenziale** *di f in $\mathbf{x}$ e viene indicata col simbolo $df(\mathbf{x})$.* Con un abuso di notazione scriveremo $df(\mathbf{x}) = < \mathbf{a},\mathbf{h} >$.

Se f è differenziabile in ogni punto di A diremo semplicemente f differenziabile in A.

Nel caso $n = 1$ il differenziale si riduce semplicemente ad ah con $a, h \in \mathbb{R}$ ed in questo caso abbiamo visto che $a = f'(x)$. Possiamo "identificare" il vettore $\mathbf{a}$ che compare nella (1.2) anche nel caso pluridimensionale?

La risposta è contenuta nel prossimo teorema insieme ad altre importanti conseguenze della differenziabilità.

■ **Teorema 1.1** - Sia $f : \mathbb{R}^n \supseteq A \to \mathbb{R}$, *$A$ aperto; se f è differenziabile in $\mathbf{x} \in A$ allora*

i) f è continua in $\mathbf{x}$;

ii) f è derivabile in $\mathbf{x}$ lungo ogni direzione; in particolare esistono tutte le derivate parziali di f in $\mathbf{x}$ e, se $\mathbf{a}$ è il vettore in (1.2), *si ha $\mathbf{a} = \nabla f(\mathbf{x})$. Inoltre vale la formula*

$$D_{\mathbf{v}} f(\mathbf{x}) = < \nabla f(\mathbf{x}), \mathbf{v} > \ . \tag{1.3}$$

Dimostrazione - *i*) Passando al limite per $\mathbf{h} \to \mathbf{0}$ nella (1.2) si ottiene $f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) \to 0$ ovvero $f(\mathbf{x}+\mathbf{h}) \to f(\mathbf{x})$.

ii) Identifichiamo prima il vettore $\mathbf{a} = (a_1, a_2, \ldots, a_n)$; scegliendo nella (1.2) $\mathbf{h} = t\mathbf{e}^j$ si ha:

$$f(\mathbf{x}+t\mathbf{e}^j) - f(\mathbf{x}) = < \mathbf{a}, t\mathbf{e}^j > + o(\|t\mathbf{e}^j\|) = t < \mathbf{a}, \mathbf{e}^j > + o(t) \ . \tag{1.4}$$

Dividendo entrambi i membri della (1.4) per t e passando al limite per $t \to 0$ si deduce che

$$f_{x_j}(\mathbf{x}) = a_j \ .$$

Dunque esistono tutte le derivate parziali in $\mathbf{x}$ e $\mathbf{a} = \nabla f(\mathbf{x})$.

Scegliendo ora nella (1.2) $\mathbf{h} = t\mathbf{v}$ dove $\mathbf{v}$ è un generico versore di $\mathbb{R}^n$ si ha:

$$f(\mathbf{x}+t\mathbf{v}) - f(\mathbf{x}) = < \mathbf{a}, t\mathbf{v} > + o(\|t\mathbf{v}\|) = t < \mathbf{a}, \mathbf{v} > + o(t) \ . \tag{1.5}$$

Dividendo entrambi i membri di (1.5) per t e passando al limite per $t \to 0$ si deduce che

$$D_{\mathbf{v}} f(\mathbf{x}) = < \mathbf{a}, \mathbf{v} > = < \nabla f(\mathbf{x}), \mathbf{v} > \ . \ \square$$

Possiamo dunque scrivere $df(\mathbf{x}) = < \nabla f(\mathbf{x}), \mathbf{h} >$; inoltre, se $f(\mathbf{x}) = x_j$ avremo $df(\mathbf{x}) = dx_j = h_j$ e perciò $df(\mathbf{x}) = < \nabla f(\mathbf{x}), d\mathbf{x} > = \sum_{j=1}^{n} f_{x_j}(\mathbf{x}) dx_j$.

La formula (1.3) permette di individuare le *direzioni di massima e minima crescita* di una funzione differenziabile. Infatti si può scrivere

$$D_{\mathbf{v}} f(\mathbf{x}) = \|\nabla f(\mathbf{x})\| \cos\beta \tag{1.6}$$

dove β è l'angolo formato dai vettori $\mathbf{v}$ e $\nabla f(\mathbf{x})$.

La (1.6) indica che $D_{\mathbf{v}} f(\mathbf{x})$ è massima quando $\beta = 0$ e quindi $\mathbf{v}_{\max} = \dfrac{\nabla f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|}$ ed è minima quando $\beta = \pi$ e quindi $\mathbf{v}_{\min} = -\dfrac{\nabla f(\mathbf{x})}{\|\nabla f(\mathbf{x})\|}$.

In conclusione

$$\max_{\|\mathbf{v}\|=1} D_{\mathbf{v}} f(\mathbf{x}) = \|\nabla f(\mathbf{x})\| \quad \text{e} \quad \min_{\|\mathbf{v}\|=1} D_{\mathbf{v}} f(\mathbf{x}) = -\|\nabla f(\mathbf{x})\| .$$

L'aspetto geometrico della differenziabilità è legato all'esistenza del piano (iperpiano se $n > 2$) tangente.

Sia f differenziabile in un punto $\mathbf{x}^0$; ponendo $\mathbf{h} = \mathbf{x} - \mathbf{x}^0$ scriviamo la (1.2) nella forma seguente:

$$f(\mathbf{x}) = f(\mathbf{x}^0) + < \nabla f(\mathbf{x}^0), \mathbf{x} - \mathbf{x}^0 > + o(\|\mathbf{x} - \mathbf{x}^0\|) . \tag{1.7}$$

La funzione $z = f(\mathbf{x}^0) + < \nabla f(\mathbf{x}^0), \mathbf{x} - \mathbf{x}^0 >$, ha come grafico un iperpiano e la (1.7) equivale ad affermare che essa è la funzione lineare (o meglio affine) che meglio approssima f in un intorno di $\mathbf{x}^0$. Tale piano si chiama *piano tangente*.

In dimensione 2 se $\mathbf{x}^0 = (x_0, y_0)$, $\mathbf{x} = (x, y)$, $z_0 = f(x_0, y_0)$ la sua equazione si scrive esplicitamente così:

$$z - z_0 - f_x(x_0, y_0)(x - x_0) - f_y(x_0, y_0)(y - y_0) = 0 . \tag{1.8}$$

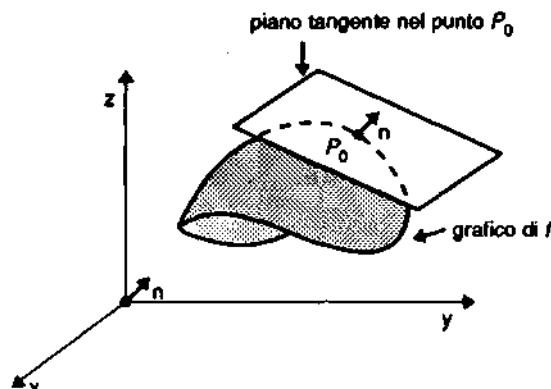



Fig. 7.4

La (1.8) indica che il vettore $\mathbf{n} = (-f_x(x_0, y_0), -f_y(x_0, y_0), 1) \in \mathbb{R}^3$ è un vettore normale al piano tangente nel punto $P_0 = (x_0, y_0, z_0)$, dunque, per definizione, normale al grafico di f nello stesso punto.

***Esempio* 1.3** - Sia

$$f(x,y) = \begin{cases} |y|^\alpha e^{-x^2/y^2} & y \neq 0 \\ 0 & y = 0 \end{cases} .$$

Esaminiamo le proprietà di f in $(0,0)$ al variare di α.

Se $\alpha \leq 0$, $f(x,y) \not\to 0$ lungo la direzione $y = x$; dunque f non è continua in $(0,0)$.

Sia $\alpha > 0$. In questo caso, essendo $0 < e^{-x^2/y^2} \leq 1$, abbiamo

$$0 \leq f(x,y) \leq |y|^\alpha$$

e quindi $f(x,y) \to 0$ per $(x,y) \to (0,0)$; cioè f è continua in $(0,0)$.

Sia $\mathbf{v} \in \mathbb{R}^2$, $\|\mathbf{v}\| = 1$; possiamo scrivere $\mathbf{v}_\theta = (\cos\theta, \sin\theta)$, $\theta \in [0, 2\pi)$.
Calcoliamo le derivate direzionali in $(0,0)$:

$$\frac{f(t\cos\theta, t\sin\theta)}{t} = \frac{|t\sin\theta|^\alpha e^{-(\operatorname{cotg}\theta)^2}}{t} \quad \text{per } \theta \neq 0, \pi$$

$$\frac{f(t,0)}{t} = 0 \quad \text{per } \theta = 0, \pi .$$

Dunque per $\theta = 0, \pi$ otteniamo $D_{\mathbf{v}_\theta} f(0,0) = \pm f_x(0,0) = 0$.

Per $\theta \neq 0, \pi$ abbiamo:
se $0 < \alpha \leq 1$, $D_{\mathbf{v}_\theta} f(0,0)$ non esiste; se $\alpha > 1$ $D_{\mathbf{v}_\theta} f(0,0) = 0$.

Sia ora $\alpha > 1$; cerchiamo i valori di α per i quali f è differenziabile in $(0,0)$.

Poiché $\nabla f(0,0) = \mathbf{0}$ si ha

$$f(x,y) - f(0,0) - f_x(0,0)x - f_y(0,0)y = |y|^\alpha e^{-x^2/y^2} .$$

Poiché

$$0 < \frac{|y|^\alpha e^{-x^2/y^2}}{\sqrt{x^2+y^2}} = \frac{|y|}{\sqrt{x^2+y^2}} |y|^{\alpha-1} e^{-x^2/y^2} \leq |y|^{\alpha-1} ,$$

essendo $\alpha > 1$, si ha che $\lim_{(x,y)\to(0,0)} |y|^\alpha e^{-x^2/y^2} = 0$, dunque f è differenziabile in $(0,0)$ e $df(0,0) = 0$ (*). □

(*) Osserviamo che $df(0,0) = 0$ è equivalente a dire che $df(0,0)$ è l'applicazione lineare nulla da $\mathbb{R}^n$ in $\mathbb{R}$ ovvero che $\nabla f(0,0) = \mathbf{0}$.

Una condizione sufficiente per la differenziabilità basata sulla conoscenza delle sole derivate parziali è contenuta nel prossimo teorema.

■ **Teorema 1.2** - *Sia* $f : \mathbb{R}^n \supseteq A \to \mathbb{R}$, *A aperto; se in un intorno di* $\mathbf{x} \in A$ *esistono tutte le derivate parziali di* f *e sono continue in* $\mathbf{x}$ *allora* f *è differenziabile in* $\mathbf{x}$.

Dimostrazione - Consideriamo il caso $n = 2$; lo studente non avrà difficoltà ad estendere la dimostrazione al caso generale.

Siano $\mathbf{x} = (x_1, x_2)$ ed $\mathbf{h} = (h_1, h_2)$; scriviamo:

$$\begin{aligned} f(\mathbf{x}+\mathbf{h}) - f(\mathbf{x}) &= f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2) = \\ &= f(x_1 + h_1, x_2 + h_2) - f(x_1 + h_1, x_2) + \\ &+ f(x_1 + h_1, x_2) - f(x_1, x_2) \; . \end{aligned} \tag{1.9}$$

Applichiamo il teorema del valor medio ai primi due termini dell'ultima somma, osservando che la sola variabile incrementata è x_2:

$$f(x_1 + h_1, x_2 + h_2) - f(x_1 + h_1, x_2) = f_{x_2}(x_1 + h_1, x_2 + \theta_2 h_2)h_2 \tag{1.10}$$

dove $\theta_2 \in (0,1)$ dipende da x_1, x_2 ed h_1, h_2.

Se usiamo ora la continuità di f_{x_2} in x_1, x_2 abbiamo

$$f_{x_2}(x_1 + h_1, x_2 + \theta_2 h_2) \to f_{x_2}(x_1, x_2) \quad \text{se} \quad (h_1, h_2) \to (0,0)$$

ovvero

$$f_{x_2}(x_1 + h_1, x_2 + \theta_2 h_2) = f_{x_2}(x_1, x_2) + \varepsilon(h_1, h_2)$$

dove $\varepsilon(h_1, h_2) \to 0$ se $(h_1, h_2) \to (0,0)$.

Si può dunque scrivere la (1.10) nella forma

$$f(x_1 + h_1, x_2 + h_2) - f(x_1 + h_1, x_2) = f_{x_2}(x_1, x_2)h_2 + h_2\varepsilon(h_1, h_2) \; . \tag{1.11}$$

Per gli ultimi due termini della (1.9), usando la definizione di $f_{x_1}(x_1, x_2)$, possiamo scrivere

$$f(x_1 + h_1, x_2) - f(x_1, x_2) = f_{x_1}(x_1, x_2)h_1 + h_1\eta(h_1) \tag{1.12}$$

dove $\eta(h_1) \to 0$ se $h_1 \to 0$.

Sostituiamo ora (1.11) e (1.12) nella (1.9); si ottiene:

$$\begin{aligned} f(x_1 + h_1, x_2 + h_2) - f(x_1, x_2) &= f_{x_1}(x_1, x_2)h_1 + f_{x_2}(x_1, x_2)h_2 + \\ &+ h_1\eta(h_1) + h_2\varepsilon(h_1, h_2) \; . \end{aligned}$$

Per concludere che f è differenziabile in (x_1, x_2) occorre mostrare che

$$h_1\eta(h_1) + h_2\varepsilon(h_1, h_2) = o(\sqrt{h_1^2 + h_2^2}\,) \;\text{ per }\; (h_1, h_2) \to (0,0) \; .$$

Si ha:

$$0 \leq \frac{|h_1\eta(h_1) + h_2\varepsilon(h_1,h_2)|}{\sqrt{h_1^2+h_2^2}} \leq \frac{|h_1|}{\sqrt{h_1^2+h_2^2}}|\eta(h_1)| + \frac{|h_2|}{\sqrt{h_1^2+h_2^2}}|\varepsilon(h_1,h_2)| \leq$$
$$\leq |\eta(h_1)| + |\varepsilon(h_1,h_2)| .$$

Poiché $\eta(h_1) \to 0$ se $h_1 \to 0$ ed $\varepsilon(h_1,h_2) \to 0$ se $(h_1,h_2) \to (0,0)$ la dimostrazione è conclusa. □

Osservazione 1.1 - Nella dimostrazione del teorema si è usata la continuità della sola derivata f_{x_2}. Nel caso generale è sufficiente richiedere la continuità di tutte le derivate parziali di f tranne una.

Osservazione 1.2 - La condizione espressa nel teorema 1.2 è solo sufficiente per la differenziabilità. Già nel caso $n = 1$ si vede che esistono funzioni differenziabili con derivata *non* continua. Un esempio è il seguente:

$$f(x) = x^2 \sin\frac{1}{x} \quad \text{se } x \neq 0 \ , \ f(0) = 0 \ .$$

Dunque la classe delle funzioni differenziabili in un aperto A contiene strettamente quella delle funzioni con derivate parziali continue in A (o, come si usa dire, delle funzioni *differenziabili con continuità*). Quest'ultima classe di funzioni si indica con il simbolo $C^1(A)$. Evidentemente $C^1(A)$ è strutturato come spazio vettoriale su $\mathbb{R}$, rispetto alla somma usuale di funzioni e al prodotto di una funzione per un numero reale.

Il teorema 1.2 può essere enunciato nella forma seguente:

se $f \in C^1(A)$, allora f è differenziabile in A.

L'operazione di differenziazione si comporta come nel caso unidimensionale rispetto alle operazioni aritmetiche; valgono infatti le seguenti semplici proprietà: siano $f, g : \mathbb{R}^n \supseteq A \to \mathbb{R}$, differenziabili in $\mathbf{x} \in A$; allora: $f+g$, fg, f/g (se $g(\mathbf{x}) \neq 0$) sono differenziabili e valgono le formule

$$d(f+g)(\mathbf{x}) = df(\mathbf{x}) + dg(\mathbf{x})$$

$$d(fg)(\mathbf{x}) = df(\mathbf{x})g(\mathbf{x}) + f(\mathbf{x})\ dg(\mathbf{x})$$

$$d(f/g)(\mathbf{x}) = \frac{g(\mathbf{x})\ df(\mathbf{x}) - f(\mathbf{x})\ dg(\mathbf{x})}{(g(\mathbf{x}))^2} \ .$$

Segnaliamo anche l'importanza pratica del differenziale come conveniente approssimazione per l'incremento di una funzione in genere non lineare. Ad esempio, nel calcolo degli errori, se una quantità q è funzione di altre quantità $m_1, m_2, \ldots, m_k$

la cui misura è affetta da un errore $dm_1, dm_2, \ldots, dm_k$, il valore di q risulterà affetto da un errore con buona approssimazione dato da

$$dq = \frac{\partial q}{\partial m_1} dm_1 + \frac{\partial q}{\partial m_2} dm_2 + \ldots + \frac{\partial q}{\partial m_k} dm_k \ .$$

Anche rispetto alla composizione di funzioni il differenziale si comporta come nel caso unidimensionale; nel caso generale però occorre far intervenire funzioni a valori vettoriali. Per questa ragione tratteremo la differenziazione delle funzioni composte nella sezione 2.

1.3 Derivate e differenziali di ordine superiore

Sia $f : \mathbb{R}^n \supseteq A \to \mathbb{R}$, A aperto. Supponiamo che, fissato il versore $\mathbf{v} \in \mathbb{R}^n$, $D_{\mathbf{v}} f$ esista in un intorno di un punto $\mathbf{x} \in A$, che denotiamo con $U(\mathbf{x})$.

Allora è definita la funzione

$$D_{\mathbf{v}} f : U(\mathbf{x}) \to \mathbb{R} \ .$$

Se ora $\mathbf{w} \in \mathbb{R}^n$ è un altro versore, ci si può chiedere se esiste $D_{\mathbf{w}} D_{\mathbf{v}} f(\mathbf{x})$, che prende il nome di *derivata seconda* di f nelle direzioni $\mathbf{v}$ e $\mathbf{w}$ (nell'ordine) e si indica con $D^2_{\mathbf{w}\,\mathbf{v}} f(\mathbf{x})$.

Nel caso particolare in cui $\mathbf{v} = \mathbf{e}^j$ e $\mathbf{w} = \mathbf{e}^k$, $D^2_{\mathbf{e}^k\mathbf{e}^j} f(\mathbf{x})$ si chiama *derivata parziale seconda rispetto* ad x_j ed x_k e si indica con uno dei simboli seguenti:

$$\frac{\partial^2 f}{\partial x_k \partial x_j}(\mathbf{x}) \ , \ f_{x_k x_j}(\mathbf{x}) \ , \ D_{x_k x_j} f(\mathbf{x}) \ , \ D^2_{kj} f(\mathbf{x}) \ , \ \partial_{x_k x_j} f(\mathbf{x}) \ .$$

Se $k \neq j$ le derivate $\dfrac{\partial^2 f}{\partial x_k \partial x_j}$ si chiamano *miste*; se $k = j$ si chiamano *pure* e il primo simbolo si semplifica in $\dfrac{\partial^2 f}{\partial x_j^2}(\mathbf{x})$.

***Esempio* 1.4** - Sia $f : \mathbf{x} \mapsto \dfrac{1}{\|\mathbf{x}\|} : \mathbb{R}^n \setminus \{\mathbf{0}\} \to \mathbb{R}$. Calcoliamo $f_{x_k x_j}(\mathbf{x})$, $\mathbf{x} \neq \mathbf{0}$. Si possono usare le ordinarie regole di derivazione, come abbiamo già osservato. Si ha:

$$f_{x_j}(\mathbf{x}) = -\frac{x_j}{\|\mathbf{x}\|^3} \ , \ f_{x_k x_j}(\mathbf{x}) = -\frac{\delta_{kj}}{\|\mathbf{x}\|^3} + 3\,\frac{x_k x_j}{\|\mathbf{x}\|^5}$$

dove δ_{jk} è il simbolo di Kronecker.

Notiamo che nell'esempio 1.4 $f_{x_k x_j} = f_{x_j x_k}$ per ogni j e k.

Sarà sempre vero, per ogni funzione due volte derivabile lungo le direzioni $\mathbf{v}$ e $\mathbf{w}$ che $D^2_{\mathbf{wv}}f = D^2_{\mathbf{vw}}f$?

Il prossimo esempio fornisce una risposta negativa.

***Esempio* 1.5** - Sia

$$f(x,y) = \begin{cases} xy\dfrac{x^2-y^2}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$$

Calcoliamo $f_{xy}(0,0)$ e $f_{yx}(0,0)$. Occorre naturalmente calcolare prima $f_x(0,y)$ e $f_y(x,0)$. Si ha:

$$f_x(0,y) = \lim_{x\to 0}\frac{f(x,y)-f(0,y)}{x} = \lim_{x\to 0} y\frac{x^2-y^2}{x^2+y^2} = -y \ ;$$

$$f_y(x,0) = \lim_{y\to 0}\frac{f(x,y)-f(x,0)}{y} = \lim_{y\to 0} x\frac{x^2-y^2}{x^2+y^2} = x \ .$$

Si conclude che

$$f_{yx}(0,0) = -1 \quad \text{e} \quad f_{xy}(0,0) = 1 \ ,$$

e le derivate miste in $(0,0)$ sono diverse.

Dunque, in generale

$$f_{x_k x_j}(\mathbf{x}) \neq f_{x_j x_k}(\mathbf{x}) \ .$$

Il prossimo teorema indica una condizione sufficiente per l'uguaglianza delle derivate seconde miste.

■ **Teorema 1.3** - (di Schwarz). *Se $f_{x_k x_j}$ e $f_{x_j x_k}$ esistono in un intorno di* $\mathbf{x}$ *e sono continue in* $\mathbf{x}$ *allora*

$$f_{x_k x_j}(\mathbf{x}) = f_{x_j x_k}(\mathbf{x}) \ .$$

Dimostrazione - Indichiamo con U l'intorno di $\mathbf{x}$ in cui esistono $f_{x_k x_j}$ e $f_{x_j x_k}$; consideriamo l'espressione

$$\Delta = f(\mathbf{x} + t\mathbf{e}^k + t\mathbf{e}^j) - f(\mathbf{x} + t\mathbf{e}^k) - f(\mathbf{x} + t\mathbf{e}^j) + f(\mathbf{x})$$

con $|t|$ abbastanza piccolo in modo che i punti $\mathbf{x} + t\mathbf{e}^k$, $\mathbf{x} + t\mathbf{e}^j$ non escano da U.

Consideriamo la funzione

$$g(\tau) = f(\mathbf{x} + t\mathbf{e}^k + \tau\mathbf{e}^j) - f(\mathbf{x} + \tau\mathbf{e}^j)$$

dove si pensa t fisso e $\tau \in [0, t]$.
Allora

$$\Delta = g(t) - g(0)$$

come facilmente si verifica. D'altra parte, variando τ, varia soltanto l'argomento di f nella direzione $\mathbf{e}^j$, ovvero varia soltanto x_j; di conseguenza

$$g'(\tau) = f_{x_j}(\mathbf{x} + t\mathbf{e}^k + \tau\mathbf{e}^j) - f_{x_j}(\mathbf{x} + \tau\mathbf{e}^j) \ .$$

Applicando il teorema di Lagrange a g, otteniamo:

$$\Delta = g(t) - g(0) = \{f_{x_j}(\mathbf{x} + t\mathbf{e}^k + \theta t\mathbf{e}^j) - f_{x_j}(\mathbf{x} + \theta t\mathbf{e}^j)\}t \tag{1.13}$$

dove θ è un opportuno numero tra 0 e 1 (dipendente da $\mathbf{x}, t$).
Introduciamo ora la funzione

$$\varphi(\tau) = f_{x_j}(\mathbf{x} + \tau\mathbf{e}^k + \theta t\mathbf{e}^j) \ ,$$

dove ancora t è fisso e $\tau \in [0, t]$. Si può scrivere, dalla (1.13):

$$\Delta = \{\varphi(t) - \varphi(0)\}t \ .$$

Questa volta, variando τ varia solo x_k e, per ipotesi, $\varphi(\tau)$ è derivabile (poiché esiste $f_{x_k x_j}$ in U). Applicando ancora il teorema di Lagrange otteniamo:

$$\Delta = \varphi'(\eta t)t^2 = f_{x_k x_j}(\mathbf{x} + \eta t\mathbf{e}^k + \theta t\mathbf{e}^j)t^2 \tag{1.14}$$

dove $\eta \in (0, 1)$ è un numero opportuno, dipendente da $\mathbf{x}$ e t.
Osserviamo ora che l'espressione di Δ è simmetrica rispetto ad x_j e x_k; ripetendo lo stesso procedimento con i ruoli di x_j ed x_k scambiati, si ottiene

$$\Delta = f_{x_j x_k}(\mathbf{x} + \eta' t\mathbf{e}^k + \theta' t\mathbf{e}^j)t^2 \ . \tag{1.15}$$

dove η' e θ' hanno le stesse proprietà di η e θ.
Dalla (1.14) e (1.15) otteniamo, dopo aver diviso per t^2:

$$f_{x_k x_j}(\mathbf{x} + \eta t\mathbf{e}^k + \theta t\mathbf{e}^j) = f_{x_j x_k}(\mathbf{x} + \eta' t\mathbf{e}^k + \theta' t\mathbf{e}^j) \ . \tag{1.16}$$

Essendo $f_{x_k x_j}$ e $f_{x_j x_k}$ continue in $\mathbf{x}$, passando al limite per $t \to 0$ nella (1.16) si deduce $f_{x_k x_j}(\mathbf{x}) = f_{x_j x_k}(\mathbf{x})$. □

Osservazione 1.3 - Il teorema vale per derivate direzionali seconde qualunque, non solo per le derivate seconde miste.

Inoltre si potrebbe dimostrare che se $f_{x_k}, f_{x_j}, f_{x_k x_j}$ esistono in un intorno di un punto $\mathbf{x}$ e $f_{x_k x_j}$ è continua in $\mathbf{x}$, allora esiste anche $f_{x_j x_k}(\mathbf{x})$ ed è uguale a $f_{x_k x_j}(\mathbf{x})$.

Oltre alle derivate seconde potremo considerare in maniera ovvia derivate di ordine superiore. Ad esempio le notazioni

$$\frac{\partial^3}{\partial x_k \partial x_j \partial x_i} f(\mathbf{x}) \; , \; f_{x_k x_j x_i}(\mathbf{x}) \; , \; D^3_{kji} f(\mathbf{x}) \; , \partial_{x_k x_j x_i} f(\mathbf{x})$$

indicano la derivata parziale terza di f rispetto a x_i, x_j, x_k nell'ordine.

Nulla di particolare v'è da aggiungere riguardo al loro calcolo.

Passiamo ora alla definizione di *differenziale secondo*.

Sia $f : \mathbb{R}^n \supseteq A \to \mathbb{R}$ differenziabile in A. Allora per ogni $\mathbf{x} \in A$ esistono le derivate parziali $f_{x_j}(\mathbf{x}), j = 1, \dots, n$.

Se queste derivate sono a loro volta differenziabili in $\mathbf{x}$ diremo che f è *due volte differenziabile* in $\mathbf{x}$ e si chiama *differenziale secondo* di f in $\mathbf{x}$ l'espressione

$$d^2 f(\mathbf{x}) := \sum_{i,j=1}^{n} f_{x_i x_j}(\mathbf{x}) dx_i dx_j \; . \tag{1.17}$$

Sottolineiamo che $d^2 f(\mathbf{x})$ è una *forma quadratica* (polinomio omogeneo di 2° grado) nelle componenti del vettore incremento $d\mathbf{x} = (dx_1, dx_2, \dots, dx_n)$. Essa giocherà un ruolo fondamentale nella teoria dell'ottimizzazione per funzioni di n variabili.

La matrice quadrata di ordine n i cui elementi sono $f_{x_i x_j}(\mathbf{x})$ si chiama *matrice Hessiana* (*) di f nel punto $\mathbf{x}$ e viene indicata col simbolo $\mathbf{H}_f(\mathbf{x})$, cioè:

$$\mathbf{H}_f(\mathbf{x}) := \begin{pmatrix} f_{x_1 x_1}(\mathbf{x}) & f_{x_1 x_2}(\mathbf{x}) & \dots & f_{x_1 x_n}(\mathbf{x}) \\ f_{x_2 x_1}(\mathbf{x}) & f_{x_2 x_2}(\mathbf{x}) & \dots & f_{x_2 x_n}(\mathbf{x}) \\ \vdots & \vdots & \dots & \vdots \\ f_{x_n x_1}(\mathbf{x}) & f_{x_n x_2}(\mathbf{x}) & \dots & f_{x_n x_n}(\mathbf{x}) \end{pmatrix}$$

Potremo allora scrivere

$$d^2 f(\mathbf{x}) = < \mathbf{H}_f(\mathbf{x}) d\mathbf{x}, \; d\mathbf{x} >$$

(*) Otto Hesse (1811–1874).

dove $\mathbf{H}_f(\mathbf{x})d\mathbf{x}$ indica il prodotto righe per colonne della matrice $\mathbf{H}_f(\mathbf{x})$ per il vettore $d\mathbf{x}$.

Se f è due volte differenziabile in $\mathbf{x}$, come immediata conseguenza abbiamo che esistono le derivate $D^2_{\mathbf{v}\mathbf{w}}f(\mathbf{x})$ per *ogni coppia* di versori $\mathbf{v}$ e $\mathbf{w} \in \mathbb{R}^n$. Inoltre vale la formula

$$D^2_{\mathbf{v}\mathbf{w}}f(\mathbf{x}) = \sum_{i,j=1}^{n} f_{x_i x_j}(\mathbf{x}) v_i w_j = <\mathbf{H}_f(\mathbf{x})\mathbf{w}, \mathbf{v}> \quad . \tag{1.18}$$

Si noti la somiglianza tra le formule (1.18) e (1.17).

La (1.18) si dimostra osservando che, essendo f differenziabile in $\mathbf{x}$, esiste $D_{\mathbf{w}}f(\mathbf{x})$ e si ha per ogni $\mathbf{x} \in A$:

$$D_{\mathbf{w}}f(\mathbf{x}) = <\nabla f(\mathbf{x}), \mathbf{w}> = \sum_{j=1}^{n} f_{x_j}(\mathbf{x}) w_j$$

dove w_j $(j = 1, \ldots, n)$ sono le componenti di $\mathbf{w}$.

Essendo le f_{x_j} differenziabili in $\mathbf{x}$ si avrà

$$D_{\mathbf{v}} f_{x_j}(\mathbf{x}) = <\nabla f_{x_j}(\mathbf{x}), \mathbf{v}> = \sum_{i=1}^{n} f_{x_i x_j}(\mathbf{x}) v_i$$

e quindi

$$D^2_{\mathbf{v}\mathbf{w}}f(\mathbf{x}) = \sum_{j=1}^{n} D_{\mathbf{v}} f_{x_j}(\mathbf{x}) w_j = \sum_{i,j=1}^{n} f_{x_i x_j}(\mathbf{x}) v_i w_j \quad \square$$

***Esempio* 1.6** - Siano $\mathbf{v} = (\cos\theta, \sin\theta)$, $\mathbf{w} = (\cos\varphi, \sin\varphi)$ e $f(x, y) = xe^{2y}$.
Allora $f_{xx} = 0$, $f_{yy} = 4xe^{2y}$, $f_{yx} = 2e^{2y}$ e

$$D^2_{\mathbf{v}\mathbf{w}}f(0,0) = 2\cos\theta\sin\varphi + 2\sin\theta\cos\varphi \ .$$

Se f è una funzione di 2 variabili differenziabile due volte in un punto $(x, y) \in A \subseteq \mathbb{R}^n$ il suo differenziale secondo in (x, y) si scrive

$$d^2 f(x,y) = f_{xx}(x,y)dx^2 + f_{xy}(x,y)dxdy + f_{yx}(x,y)dxdy + f_{yy}(x,y)dy^2 \ .$$

Il fatto che f sia due volte differenziabile non implica la continuità di f_{xy} e f_{yx}: non si può dunque applicare il teorema 1.3 per concludere che $f_{xy}(x, y) = f_{yx}(x, y)$. Che ciò sia vero segue però dal seguente teorema:

■ **Teorema 1.4** - *Se f è due volte differenziabile in* $\mathbf{x}$ *l'ordine di derivazione nelle derivate miste è invertibile.*

Dimostrazione - Facciamo vedere che $f_{x_k x_j}(\mathbf{x}) = f_{x_j x_k}(\mathbf{x})$. Il ragionamento è simile a quello della dimostrazione del teorema 1.3. Sia dunque Δ come in quella dimostrazione;

usando il teorema di Lagrange possiamo ancora scrivere, poiché le derivate prime esistono in un intorno di $\mathbf{x}$:

$$\Delta = \{f_{x_j}(\mathbf{x}+t\,\mathbf{e}^k+\theta t\,\mathbf{e}^j) - f_{x_j}(\mathbf{x}+\theta t\,\mathbf{e}^j)\}t$$

per $\theta \in (0,1)$ opportuno.

Non possiamo ora utilizzare il teorema di Lagrange poiché sappiamo che le derivate seconde esistono in $\mathbf{x}$ soltanto.

Sappiamo però che f_{x_j} è differenziabile in $\mathbf{x}$, essendo f due volte differenziabile. Possiamo allora scrivere:

$$f_{x_j}(\mathbf{x}+t\,\mathbf{e}^k+\theta t\,\mathbf{e}^j) = f_{x_j}(\mathbf{x}) + f_{x_k x_j}(\mathbf{x})t + f_{x_j x_j}(\mathbf{x})\theta t + \eta(\mathbf{x},t)$$

$$f_{x_j}(\mathbf{x}+\theta t\,\mathbf{e}^j) = f_{x_j}(\mathbf{x}) + f_{x_j x_j}(\mathbf{x})\;\theta t + \eta_1(\mathbf{x},t)$$

dove η ed η_1 sono infinitesimi di ordine superiore a t per $t \to 0$.

Allora:

$$\Delta = \{f_{x_k x_j}(\mathbf{x})t + \eta(\mathbf{x},t) - \eta_1(\mathbf{x},t)\}t\;. \tag{1.19}$$

Analogamente, sfruttando la simmetria di Δ rispetto ad x_k ed x_j, si ha:

$$\Delta = \{f_{x_j x_k}(\mathbf{x})t + \overline{\eta}(\mathbf{x},t) - \overline{\eta}_1(\mathbf{x},t)\}t \tag{1.20}$$

dove ancora $\overline{\eta}$ e $\overline{\eta_1}$ sono infinitesimi di ordine superiore a t per $t \to 0$.

Dalla (1.19) si deduce

$$\lim_{t\to 0}\frac{\Delta}{t^2} = f_{x_k x_j}(\mathbf{x})\;,$$

mentre dalla (1.20) si ha:

$$\lim_{t\to 0}\frac{\Delta}{t^2} = f_{x_j x_k}(\mathbf{x})\;.$$

Di conseguenza

$$f_{x_j x_k}(\mathbf{x}) = f_{x_k x_j}(\mathbf{x})\;. \;\square$$

In base al teorema 1.4 si può scrivere, per una funzione due volte differenziabile di due variabili

$$d^2 f(x,y) = f_{xx}(x,y)dx^2 + 2f_{xy}(x,y)dxdy + f_{yy}(x,y)dy^2\;. \tag{1.21}$$

È evidente l'analogia della (1.21) con la formula per il quadrato di un binomio. Questa analogia può essere spinta più in profondità e, come vedremo, generalizzata.

A tale scopo, data la solita funzione $f : \mathbb{R}^n \supseteq A \to \mathbb{R}$, riguardiamo le derivate $\dfrac{\partial f}{\partial x_j}$ come effetto dell'azione dell'*operatore* $\dfrac{\partial}{\partial x_j}$ sulla funzione f:

$$f \longrightarrow \boxed{\frac{\partial}{\partial x_j}} \longrightarrow \frac{\partial f}{\partial x_j}\,.$$

Incidentalmente osserviamo che l'azione dell'operatore $\dfrac{\partial}{\partial x_j}$ può essere automatizzata per mezzo dei cosiddetti "linguaggi formali" che l'informatica ci mette oggi a disposizione.

Così come si può definire la composizione di funzioni, si può definire la composizione di operatori come $\dfrac{\partial}{\partial x_j}$, che vengono chiamati *operatori differenziali.*

In questo modo $\dfrac{\partial^2}{\partial x_k \partial x_j}$ è interpretabile come composizione di $\dfrac{\partial}{\partial x_j}$ e $\dfrac{\partial}{\partial x_k}$, nell'ordine, ovvero come "prodotto di $\dfrac{\partial}{\partial x_j}$ e $\dfrac{\partial}{\partial x_k}$":

$$f \longrightarrow \boxed{\frac{\partial}{\partial x_j}} \longrightarrow \frac{\partial f}{\partial x_j} \longrightarrow \boxed{\frac{\partial}{\partial x_k}} \rightarrow \frac{\partial^2 f}{\partial x_k \partial x_j}\;;$$

$\dfrac{\partial^2}{\partial x_k^2}$ è interpretabile come composizione di $\dfrac{\partial}{\partial x_k}$ con sé stesso, ovvero come il "quadrato" di $\dfrac{\partial}{\partial x_k}$ che possiamo indicare con $\left(\dfrac{\partial}{\partial x_k}\right)^2$.

Con questa convenzione di scrittura la (1.21) diventa

$$d^2 f = \left(\frac{\partial}{\partial x}dx + \frac{\partial}{\partial y}dy\right)^2 f \quad \text{(si legge: quadrato formale di ...)}$$

dove il quadrato agisce nel modo usuale sui differenziali dx e dy.

In generale, per una funzione 2 volte differenziabile in $\mathbf{x}$, si avrà:

$$d^2 f = \left(\frac{\partial}{\partial x_1}dx_1 + \frac{\partial}{\partial x_2}dx_2 + \ldots + \frac{\partial}{\partial x_n}dx_n\right)^2 f\,. \tag{1.22}$$

I differenziali di ordine superiore al secondo si possono definire facilmente. Sia $k > 2$ e sia f una funzione dotata di *tutte le derivate parziali di ordine* $k-1$ in un intorno di $\mathbf{x}$.

Se ognuna di queste derivate è differenziabile in $\mathbf{x}$ allora si dice che f è k *volte differenziabile* in $\mathbf{x}$. Si noti che, in base ai teoremi 1.3 e 1.4, se f è k volte differenziabile in $\mathbf{x}$, esistono in un intorno di $\mathbf{x}$ tutte le derivate parziali dall'ordine 1 all'ordine $k-1$ e sono ivi differenziabili; inoltre per ognuna di esse vale il teorema relativo all'inversione dell'ordine di derivazione.

Il differenziale di ordine k di f in $\mathbf{x}$ è assegnato dalla formula seguente:

$$d^k f(\mathbf{x}) := \left(\frac{\partial}{\partial x_1}dx_1 + \frac{\partial}{\partial x_2}dx_2 + \ldots + \frac{\partial}{\partial x_k}dx_k\right)^k =$$
$$= \sum_{i_1,i_2,\ldots,i_k=1}^{n} \frac{\partial^k f}{\partial x_{i_1}\partial x_{i_2}\ldots\partial x_{i_k}}(\mathbf{x})dx_{i_1}dx_{i_2}\ldots dx_{i_k}\,.$$

Si vede che $d^k f(\mathbf{x})$ è un polinomio omogeneo di grado k nelle componenti di $d\mathbf{x}$.

Importanti classi di funzioni sono $C^k(A)$ e $C^\infty(A)$; $C^k(A)$ indica la classe delle funzioni k volte differenziabili con continuità, ovvero dotate di derivate parziali fino all'ordine k incluso, continue in A; esse sono dunque k volte differenziabili in base al teorema 1.3 (applicato alle derivate di ordine $k-1$); $C^\infty(A)$ indica l'insieme delle funzioni infinite volte differenziabili con continuità.

Concludiamo il paragrafo con una serie di osservazioni riguardanti gli operatori differenziali.

Consideriamo $\dfrac{\partial}{\partial x_j}$; le proprietà dell'operazione di derivata implicano immediatamente che, $\forall\alpha,\beta\in\mathbb{R}$, $\forall f,g$ derivabili:

$$\frac{\partial}{\partial x_j}(\alpha f + \beta g) = \alpha\frac{\partial f}{\partial x_j} + \beta\frac{\partial g}{\partial x_j}$$

ovvero che l'*operatore* $\dfrac{\partial}{\partial x_j}$ è *lineare*.

Se inoltre consideriamo $\dfrac{\partial}{\partial x_k}$, in base al teorema di Schwarz, per funzioni di classe $C^2(A)$ abbiamo $\dfrac{\partial^2 f}{\partial x_k\partial x_j} = \dfrac{\partial^2 f}{\partial x_j\partial x_k}$, formula che possiamo interpretare come *commutatività* degli operatori $\dfrac{\partial}{\partial x_i}$ e $\dfrac{\partial}{\partial x_k}$.

Anche le derivate direzionali possono essere interpretate come operatori differenziali con le stesse proprietà di linearità. Fissato $\mathbf{v}\in\mathbb{R}^n$, $\|\mathbf{v}\|=1$, l'operatore $D_\mathbf{v}$ associa ad f la derivata $D_\mathbf{v}f$.

Un altro importante operatore differenziale è il seguente:

$$\Delta : f \mapsto \Delta f := \frac{\partial^2 f}{\partial x_1^2} + \frac{\partial^2 f}{\partial x_2^2} + \ldots + \frac{\partial^2 f}{\partial x_n^2} \quad \textit{(operatore di Laplace)}\,.$$

Le funzioni $f\in C^2(A)$ tali che $\Delta f = 0$ in A si dicono *armoniche*. Ad esempio $f(x,y) = x^2 - y^2$ è armonica in $\mathbb{R}^2$. Queste funzioni hanno un ruolo particolarmente importante in fisica-matematica.

1.4 Formula di Taylor

Per una funzione differenziabile in un punto $\mathbf{x}$, $df(\mathbf{x})$ costituisce la migliore approssimazione lineare in un intorno di $\mathbf{x}$; approssimazioni locali più accurate si ottengono facendo intervenire i differenziali di ordine superiore, come nel caso unidimensionale. Ciò che si ottiene è l'estensione della formula di Taylor. Il prossimo teorema estende tale formula con il resto nella forma di Lagrange.

■ **Teorema 1.5** - *Sia* $f : \mathbb{R}^n \supseteq A \to \mathbb{R}$ *e supponiamo che il segmento chiuso* $[\mathbf{x}, \mathbf{x}+d\mathbf{x}]$ *sia contenuto in* A.
Se f *è differenziabile* $k-1$ *volte nei punti del segmento chiuso e* k *volte differenziabile in quelli del segmento aperto, allora esiste* $\theta \in (0,1)$ *tale che*

$$f(\mathbf{x}+d\mathbf{x}) - f(\mathbf{x}) = df(\mathbf{x}) + \frac{1}{2}d^2 f(\mathbf{x}) + \ldots + \frac{1}{(k-1)!}d^{k-1}f(\mathbf{x}) + \frac{1}{k!}d^k f(\mathbf{x}+\theta d\mathbf{x}) \,. \tag{1.23}$$

La (1.23) si chiama Formula di Taylor di ordine k con centro nel punto $\mathbf{x}$ (*).
Per la dimostrazione facciamo uso della seguente proposizione, che risulta essere un caso particolare del teorema 2.1, presentato nella sezione 2.

Proposizione 1.6 - *Sia* $f : \mathbb{R}^n \supseteq A \to \mathbb{R}$, *differenziabile in* $\mathbf{x} \in A$; *siano inoltre* $x_1 = x_1(t), x_2 = x_2(t), \ldots, x_n = x_n(t)$ *funzioni da* $(a,b) \to A$, *differenziabili in* (a,b) *e tali che* $\mathbf{x} = (x_1(t_0), x_2(t_0), \ldots, x_n(t_0))$. *Allora la funzione composta*

$$F(t) = f(x_1(t), x_2(t), \ldots, x_n(t))$$

è differenziabile in t_0 *e vale la formula*

$$F'(t_0) = \frac{\partial f}{\partial x_1}(\mathbf{x}) \cdot x_1'(t_0) + \frac{\partial f}{\partial x_2}(\mathbf{x}) \cdot x_2'(t_0) + \ldots + \frac{\partial f}{\partial x_n}(\mathbf{x}) \cdot x_n'(t_0) \,. \tag{1.24}$$

Dimostrazione del teorema 1.5 - L'idea è di ricondursi al caso unidimensionale introducendo la funzione $g(t) = f(\mathbf{x}+t d\mathbf{x})$ ed osservando che:

i) $f(\mathbf{x}+d\mathbf{x}) - f(\mathbf{x}) = g(1) - g(0)$;

ii) in base alla proposizione 1.6, con $x_j(t) = x_j + t dx_j$ $(j = 1, \ldots, n)$, g risulta $k-1$ volte differenziabile in $[0,1]$ e k volte differenziabile in $(0,1)$.

(*) Si chiama ancora formula di Mac Laurin se $\mathbf{x} = \mathbf{0}$.

iii) Inoltre poiché $x_j'(t) = dx_j$ per ogni $j = 1, \ldots, n$ si ha, in base alla formula (1.24)

$$g'(t) = \frac{\partial f}{\partial x_1}(\mathbf{x}+td\mathbf{x})dx_1 + \frac{\partial f}{\partial x_2}(\mathbf{x}+td\mathbf{x})dx_2 + \ldots + \frac{\partial f}{\partial x_n}(\mathbf{x}+td\mathbf{x})dx_n$$

cosicché $g'(0) = df(\mathbf{x})$.
Analogamente si ricava che

$$g''(0) = d^2 f(\mathbf{x}), \ldots, g^{(k-1)}(0) = d^{k-1} f(\mathbf{x})$$

ed infine

$$g^{(k)}(t) = d^k f(\mathbf{x}+td\mathbf{x}) \; .$$

Applicando la formula di Taylor a g per l'intervallo $[0, 1]$ possiamo scrivere:

$$g(1) - g(0) = g'(0) + \frac{1}{2}g''(0) + \ldots + \frac{1}{(k-1)!}g^{(k-1)}(0) + \frac{1}{k!}g^{(k)}(\theta) \qquad (1.25)$$

dove $\theta \in (0, 1)$, opportuno.
Tenendo conto di *i)*, *ii)* e *iii)*, la (1.25) coincide esattamente con la (1.23). □

Il caso particolare $k = 1$ nella (1.23) è l'estensione al caso di funzioni reali di n variabili del teorema del valor medio di Lagrange:

$$f(\mathbf{x}+d\mathbf{x}) - f(\mathbf{x}) = df(\mathbf{x}+\theta d\mathbf{x}) \quad (\theta \in (0,1) \text{ opportuno}) \; . \qquad (1.26)$$

Come già nel caso unidimensionale tramite la (1.26) possiamo caratterizzare le funzioni costanti di un *aperto connesso.*

Proposizione 1.7 - *Sia A aperto connesso e $df(\mathbf{x}) = 0$ $\forall \mathbf{x} \in A$. Allora f è costante.*
La dimostrazione della proposizione 1.7 si ricava facilmente ricordando che A aperto connesso è connesso per segmenti. Lasciamo i dettagli all'allievo.

Come vedremo nella sezione 2, la formula di Taylor con il resto di Lagrange *non* è estendibile al caso di funzioni vettoriali e quindi neppure il teorema del valor medio.
Si potrà invece estendere anche a queste funzioni la formula di Taylor con il resto di Peano, oggetto del prossimo teorema.

■ **Teorema 1.8** - *Sia $f : \mathbb{R}^n \supseteq A \to \mathbb{R}$, k volte differenziabile in $\mathbf{x}$. Allora*

$$f(\mathbf{x}+d\mathbf{x}) - f(\mathbf{x}) = df(\mathbf{x}) + \frac{1}{2}d^2 f(\mathbf{x}) + \ldots + \frac{1}{k!}d^k f(\mathbf{x}) + o(\|d\mathbf{x}\|^k) \qquad (1.27)$$

per $\|d\mathbf{x}\| \to 0$.

Dimostrazione - Ne diamo solo l'idea nel caso $n = 2$, $k = 2$.
Poniamo $d\mathbf{x} = (h, k), \mathbf{x} = (x, y)$ e

$$g(h,k) = f(x+h, y+k) - f(x,y) - f_x(x,y)h - f_y(x,y)k -$$
$$-\frac{1}{2}\{f_{xx}(x,y)h^2 + 2f_{xy}(x,y)hk + f_{yy}(x,y)k^2\} .$$

Occorre dimostrare che $g(h,k)/(h^2+k^2) \to 0$ quando $h^2 + k^2 \to 0$.
Usiamo il teorema del valor medio, osservando che $g(0,0) = 0$:

$$g(h,k) = g(h,k) - g(0,0) = g_h(\theta h, \theta k)h + g_k(\theta h, \theta k)k \tag{1.28}$$

con θ opportuno tra 0 e 1, $\theta = \theta(h,k)$. Ora abbiamo:

$$g_h(h,k) = f_x(x+h, y+k) - f_x(x,y) - f_{xx}(x,y)h - f_{xy}(x,y)k$$

e quindi, essendo f_x differenziabile in (x,y), si ha $g_h(h,k) = o(\sqrt{h^2+k^2})$.
Analogamente

$$g_k(h,k) = f_y(x+h, y+k) - f_y(x,y) - f_{yy}(x,y)k - f_{xy}(x,y)h = o(\sqrt{h^2+k^2})$$

in virtù della differenziabilità di f_y.
Sostituendo nella (1.28) otteniamo, essendo $0 < \theta = \theta(h,k) < 1$:

$$|g(h,k)| = |(h+k) \cdot o(\sqrt{(\theta h)^2 + (\theta k)^2})| \leq (|h| + |k|) \cdot |o(\sqrt{h^2+k^2})|$$

da cui $|g(h,k)|/(h^2+k^2) \leq |o(\sqrt{h^2+k^2})|/\sqrt{h^2+k^2}$ e quindi la tesi. □

1.5 Funzioni omogenee; funzioni convesse e concave

Presentiamo in questo paragrafo due particolari classi di funzioni spesso impiegate nelle applicazioni.

La prima è quella delle funzioni *positivamente omogenee di grado* $\mu, \mu \in \mathbb{R}$. Queste funzioni sono definite in *coni* con vertice in $\mathbf{0}$, vale a dire che, se sono definite in un punto $\mathbf{x}$, allora sono definite su tutta la semiretta $\rho\mathbf{x}$ per ogni $\rho > 0$. Il cono non è necessariamente convesso e, naturalmente, può coincidere con tutto $\mathbb{R}^n$.

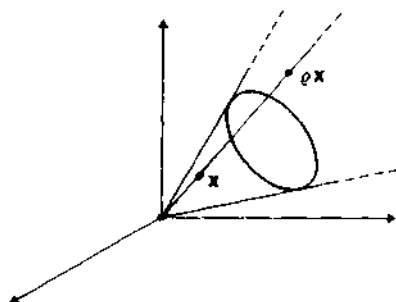



Fig. 7.5 Un cono convesso in $\mathbb{R}^3$.

Definizione 1.3 - *Sia $C \subseteq \mathbb{R}^n$ un cono e $f : C \to \mathbb{R}$; f si dice positivamente omogenea di grado $\mu, \mu \in \mathbb{R}$, se $\forall \mathbf{x} \in C$ e $\forall \rho > 0$ risulta*

$$f(\rho \mathbf{x}) = \rho^{\mu} f(\mathbf{x}) . \tag{1.29}$$

Esempi

1.7. Qualunque *polinomio omogeneo* di grado $\mu, \mu \in \mathbb{N}$, in k variabili (ovvero ogni termine ha grado μ) è omogeneo nel senso della definizione 1.3, dello stesso grado.

Ad esempio

$$f(\mathbf{x}) = a_1 x_1 + a_2 x_2 + \ldots + a_k x_k \quad (\mu = 1)$$

$$f(\mathbf{x}) = \sum_{i,j=1}^{k} a_{ij} x_i x_j . \qquad (\mu = 2)$$

La prima funzione si dice anche linearmente omogenea; la seconda si chiama *forma quadratica* e si può scrivere nel modo seguente

$$\sum_{i,j=1}^{k} a_{ij} x_i x_j = < \mathbf{A}\,\mathbf{x}, \mathbf{x} >$$

dove $\mathbf{A}$ è la matrice quadrata di ordine k i cui elementi sono a_{ij} e $\mathbf{x} = (x_1, x_2, \ldots, x_k)$.

In particolare il differenziale secondo è una forma quadratica (nelle componenti di $d\mathbf{x}$). Per questa ragione le forme quadratiche avranno un ruolo speciale nella teoria dell'ottimizzazione per funzioni di più variabili.

1.8. La norma di un vettore $\mathbf{x} \in \mathbb{R}^n$ è positivamente omogenea di grado 1. Infatti, se $\rho > 0$

$$||\rho \mathbf{x}|| = \rho ||\mathbf{x}|| .$$

1.9. Siano $\mathbf{x}$ ed $\mathbf{y}$ vettori non nulli di $\mathbb{R}^3$ e sia f la funzione che associa ad ogni coppia $(\mathbf{x}, \mathbf{y})$ il coseno dell'angolo tra $\mathbf{x}$ ed $\mathbf{y}$, in formule:

$$f(\mathbf{x}, \mathbf{y}) = \frac{< \mathbf{x}, \mathbf{y} >}{||\mathbf{x}|| \cdot ||\mathbf{y}||} .$$

Allora f è positivamente omogenea di *grado zero* ($\mu = 0$).

Il seguente teorema caratterizza le funzioni positivamente omogenee differenziabili in coni aperti.

■ **Teorema 1.9** - (di Eulero) *Sia $C \subseteq \mathbb{R}^n$ un cono aperto e $f : C \to \mathbb{R}$ differenziabile in C. Allora f è positivamente omogenea di grado μ se e solo se, $\forall \mathbf{x} \in C$ risulta:*

$$< \mathbf{x}, \nabla f(\mathbf{x}) >= \mu f(\mathbf{x}) . \tag{1.30}$$

Dimostrazione - Se f è positivamente omogenea di grado μ, allora, derivando l'identità (1.29) rispetto a ρ e usando la proposizione 1.6, si ottiene

$$\frac{\partial f}{\partial x_1}(\rho\mathbf{x})x_1 + \frac{\partial f}{\partial x_2}(\rho\mathbf{x})x_2 + \ldots + \frac{\partial f}{\partial x_n}(\rho\mathbf{x})x_n = \mu\rho^{\mu-1}f(\mathbf{x}) .$$

Ponendo $\rho = 1$ si deduce la (1.30).
Viceversa, sia valida la (1.30); poniamo

$$\varphi(\rho) = f(\rho\mathbf{x})/\rho^{\mu}$$

e facciamo vedere che $\varphi'(\rho) = 0$ in $\mathbb{R}_+$. Si ha:

$$\varphi'(\rho) = \frac{<\mathbf{x}, \nabla f(\rho\mathbf{x})> \rho^{\mu} - \mu\rho^{\mu-1}f(\rho\mathbf{x})}{\rho^{2\mu}} = \frac{1}{\rho^{\mu+1}}\{<\rho\mathbf{x}, \nabla f(\rho\mathbf{x})> -\mu f(\rho\mathbf{x})\} = 0$$

per la (1.30), con $\rho\mathbf{x}$ al posto di $\mathbf{x}$. Dunque $\varphi(\rho)$ è costante in $\mathbb{R}_+$ e quindi $\varphi(\rho) = \varphi(1)$ ovvero $f(\rho\mathbf{x}) = \rho^{\mu}f(\mathbf{x})$. □

Estendiamo ora al caso di funzioni reali di più variabili la nozione di funzione *convessa* e *concava*, che, del resto, è formalmente identica a quella del caso unidimensionale.
Sia $f : \mathbb{R}^n \supseteq A \to \mathbb{R}$. Indichiamo con $Epi(f)$ l'*epigrafico* di f, cioè

$$Epi(f) := \{(\mathbf{x}, z) \in \mathbb{R}^{n+1} : z \geq f(\mathbf{x}),\ \mathbf{x} \in A\} .$$

Definizione 1.4 - *Sia f definita su un insieme convesso $A \subseteq \mathbb{R}^n$; f si dice convessa in A se $Epi(f)$ è convesso in $\mathbb{R}^{n+1}$; f si dice concava se $-f$ è convessa.*

Alternativamente: f è convessa in A se per ogni coppia di punti $\mathbf{x}, \mathbf{y} \in A$ e per ogni $t \in (0,1)$ risulta

$$f(t\mathbf{y} + (1-t)\mathbf{x}) \leq tf(\mathbf{y}) + (1-t)f(\mathbf{x}) . \tag{1.31}$$

La (1.31) significa che i punti del segmento di estremi $(\mathbf{x}, f(\mathbf{x}))$ e $(\mathbf{y}, f(\mathbf{y}))$ stanno al di sopra del grafico di f o, al più, sul grafico di f (vedi fig 7.6).
Se la disuguaglianza (1.30) vale in senso stretto, f si dice *strettamente convessa*.

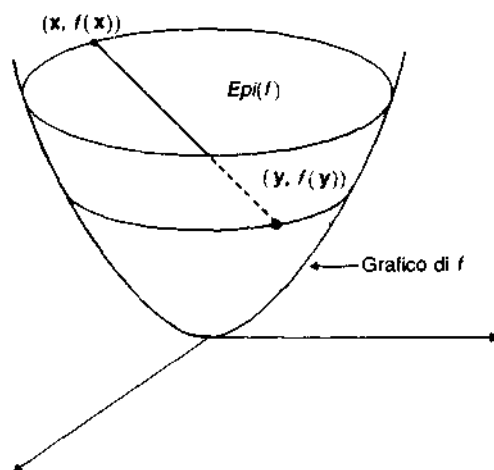



Fig. 7.6 Funzione strettamente convessa.

Per le funzioni concave (strettamente concave) la (1.31) vale con $\geq$ ($>$) anziché $\leq$ ($<$).

Ci riferiremo d'ora in poi alle funzioni convesse.

Esempi

1.10. $f(x, y) = x^2 + y^2$ è strettamente convessa in $\mathbb{R}^2$;

$f(\mathbf{x}) = \sum_{j=1}^{n} x_j^2 = ||\mathbf{x}||^2$ è strettamente convessa in $\mathbb{R}^n$.

1.11. $f(\mathbf{x}) = \sqrt{x^2 + y^2}$ è strettamente convessa in $\mathbb{R}^2$.

$f(\mathbf{x}) = ||\mathbf{x}||$ è strettamente convessa in $\mathbb{R}^n$.

1.12. Se f è convessa in A, allora anche $f^\alpha (\alpha > 1, f > 0)$, e^f sono convesse. Si veda in riferimento a questo esempio l'esercizio 23 della presente sezione.

Come per le funzioni convesse di una variabile, quelle di più variabili hanno, per il solo fatto di essere convesse, notevoli proprietà di regolarità, come risulta dal seguente teorema che ci limitiamo ad enunciare.

■ **Teorema 1.10** - *Sia f convessa in A, aperto convesso di $\mathbb{R}^n$. Allora:*

i) f è continua in A;

ii) in ogni punto di A, f ammette derivate parziali destre e sinistre;

iii) nei punti in cui esistono tutte le derivate parziali, f è differenziabile.

Se una funzione convessa di una variabile è derivabile in un punto x_0, la tangente

al grafico nel punto di ascissa x_0 sta sotto (o almeno non sopra) il grafico stesso. L'equivalente di questa affermazione per funzioni convesse di due variabili sarebbe: se $f = f(x, y)$ è *differenziabile* in (x_0, y_0) e convessa allora il piano tangente sta sotto (o almeno non sopra) il grafico di f.

L'affermazione è vera, come mostra il seguente teorema.

■ **Teorema 1.11** - *Sia f differenziabile in A aperto convesso di $\mathbb{R}^n$; f è convessa in A (strettamente convessa) se e solo se per ogni* $\mathbf{x}, \mathbf{y} \in A$*:*

$$f(\mathbf{y}) \geq f(\mathbf{x}) + df(\mathbf{x}) \quad (f(\mathbf{y}) > f(\mathbf{x}) + df(\mathbf{x}), \mathbf{x} \neq \mathbf{y}) . \tag{1.32}$$

La (1.32), scritta esplicitamente per una funzione di due variabili, assume la forma seguente:

$$f(y_1, y_2) \geq f(x_1, x_2) + f_{x_1}(x_1, x_2)(y_1 - x_1) + f_{x_2}(x_1, x_2)(y_2 - x_2)$$

cioè in corrispondenza al punto (y_1, y_2) la quota sul grafico di f è maggiore o uguale a quella sul piano tangente (nel punto$(x_1, x_2, f(x_1, x_2))$).

Dimostrazione - Sia f convessa. Scriviamo la (1.31) nella forma

$$\frac{f(t(\mathbf{y} - \mathbf{x}) + \mathbf{x}) - f(\mathbf{x})}{t} \leq f(\mathbf{y}) - f(\mathbf{x}) .$$

Poiché f è differenziabile, il primo membro della disuguaglianza, per $t \to 0$, tende a $< \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} >= df(\mathbf{x})$, da cui la (1.32).

Viceversa, se vale la (1.32) posto $\mathbf{z}_t = t\mathbf{y} + (1 - t)\mathbf{x}$ abbiamo:

$$f(\mathbf{y}) \geq f(\mathbf{z}_t) + < \nabla f(\mathbf{z}_t), \mathbf{y} - \mathbf{z}_t > = f(\mathbf{z}_t) + < \nabla f(\mathbf{z}_t), \mathbf{y} - \mathbf{x} > (1 - t)$$

$$f(\mathbf{x}) \geq f(\mathbf{z}_t) + < \nabla f(\mathbf{z}_t), \mathbf{x} - \mathbf{z}_t > = f(\mathbf{z}_t) + < \nabla f(\mathbf{z}_t), \mathbf{y} - \mathbf{x} > (-t) .$$

Moltiplicando la prima disuguaglianza per t, la seconda per $(1 - t)$ e sommando membro a membro si ottiene

$$t f(\mathbf{y}) + (1 - t) f(\mathbf{x}) \geq f(\mathbf{z}_t)$$

che è la condizione di convessità. □

Concludiamo questo paragrafo con una condizione "del secondo ordine" sufficiente per la stretta convessità di funzioni due volte differenziabili, analoga al segno della derivata seconda per funzioni di una variabile. Essa è una facile conseguenza della formula di Taylor e del teorema 1.11.

■ **Teorema 1.12** - *Sia f due volte differenziabile in A, aperto convesso di $\mathbb{R}^n$: f è strettamente convessa in A se per ogni* $\mathbf{x} \in A$, $d^2 f(\mathbf{x})$ *è una forma quadratica definita positiva (ovvero per ogni* $d\mathbf{x} \neq \mathbf{0}$, $d^2 f(\mathbf{x}) > 0$*).*

Dimostrazione - Sia $\mathbf{y} = \mathbf{x} + d\mathbf{x}$ con $\mathbf{y}, \mathbf{x} \in A$; se $d^2 f(\mathbf{z}) > 0 \ \forall \mathbf{z} \in A$ e $\forall d\mathbf{x} \neq \mathbf{0}$ si ha, per un θ opportuno, $0 < \theta < 1$:

$$f(\mathbf{y}) = f(\mathbf{x}) + df(\mathbf{x}) + \frac{1}{2} d^2 f(\mathbf{x} + \theta d\mathbf{x}) > f(\mathbf{x}) + df(\mathbf{x})$$

da cui la stretta convessità in base al teorema 1.11. □

Osserviamo che per funzioni di due variabili, due volte differenziabili, si ha

$$d^2 f(x_1, x_2) = f_{x_1 x_1}(x_1, x_2) dx_1^2 + 2 f_{x_1 x_2}(x_1, x_2) dx_1 dx_2 + f_{x_2 x_2}(x_1, x_2) dx_2^2 \, .$$

Posto $a = f_{x_1 x_1}(x_1, x_2)$, $b = f_{x_1 x_2}(x_1, x_2)$, $c = f_{x_2 x_2}(x_1, x_2)$, il trinomio $a dx_1^2 + 2b dx_1 dx_2 + c dx_2^2$ è sempre positivo qualunque sia $(dx_1, dx_2) \neq (0,0)$ se e solo se $a > 0$ e $b^2 - ac < 0$ ovvero se

$$\operatorname{tr} \mathbf{H}_f(x_1, x_2) = \Delta f(x_1, x_2) > 0 \quad \text{e} \quad \det \mathbf{H}_f(x_1, x_2) > 0 \, , \qquad (1.33)$$

dove $\mathbf{H}_f$ indica la matrice Hessiana di f.

Dunque se $z = f(x_1, x_2)$ è due volte differenziabile, allora è strettamente convessa se valgono le (1.33).

Esempi

1.13. Sia $f(x_1, x_2) = e^{x_1 + x_2^2}$; si ha

$$f_{x_1 x_1} = e^{x_1 + x_2^2} \ , \ f_{x_1 x_2} = 2x_2 e^{x_1 + x_2^2} \ , \ f_{x_2 x_2} = (2 + 4x_2^2) e^{x_1 + x_2^2} \, .$$

Dunque

$$\Delta f = e^{x_1 + x_2^2}(3 + 4x_2^2) > 0 \quad \forall x_1, x_2 \in \mathbb{R}^2$$

e

$$\det \mathbf{H}_f = e^{2(x_1 + x_2^2)}(2 + 4x_2^2) - 4x_2^2 e^{2(x_1 + x_2^2)} = 2e^{2(x_1 + x_2^2)} > 0$$

per cui f è strettamente convessa in $\mathbb{R}^2$.

Esercizi

1. Per le seguenti funzioni

$$a) \ f(x,y) = \begin{cases} |y|^\alpha \dfrac{\cos x}{\sqrt{x^2 + y^2}} & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}$$

$$b)\ f(x,y,z)=\begin{cases} z^4\dfrac{(x^2+y^2)^\alpha}{(x^2+y^2+z^2)} & (x,y,z)\neq(0,0,0)\\ 0 & (x,y,z)=(0,0,0)\end{cases}$$

studiare, al variare del parametro reale positivo α, continuità, derivabilità, differenziabilità nell'origine.

2. Posto $r=\|\mathbf{x}\|=\sqrt{x_1^2+x_2^2+\ldots+x_n^2}$, $n>1$, calcolare, nei punti $\mathbf{x}\neq\mathbf{0}$, $\Delta(r^{-1})$. Per quale valore di n si ha $\Delta(r^{-1})=0$?

3. Dimostrare le seguenti proprietà della derivata direzionale:

a) $D_\mathbf{v}(\alpha g+\beta f)=\alpha D_\mathbf{v}g+\beta D_\mathbf{v}f \quad \forall\alpha,\beta\in\mathbb{R},\ \forall\mathbf{v}\in\mathbb{R}^n$.

b) $D_\mathbf{v}(fg)=fD_\mathbf{v}g+gD_\mathbf{v}f$

c) $D_\mathbf{v}(f/g)=\dfrac{gD_\mathbf{v}f-fD_\mathbf{v}g}{g^2}$.

È vera anche la seguente proprietà?

d) $D_{\alpha\mathbf{v}+\beta\mathbf{w}}f=\alpha D_\mathbf{v}f+\beta D_\mathbf{w}f$.

4. Dimostrare le seguenti proprietà del gradiente:

a) $\nabla(\alpha f+\beta g)=\alpha\nabla f+\beta\nabla g$

b) $\nabla(fg)=f\nabla g+g\nabla f$

c) $\nabla(f/g)=\dfrac{g\nabla f-f\nabla g}{g^2}$.

Valgono analoghe proprietà per il differenziale?

5. Verificare che per la funzione dell'esempio 1.2 non vale la formula (1.3).

6. Siano f,g armoniche in $\mathbb{R}^n$ (cioè $\Delta f=\Delta g=0$). Quando fg è armonica?

7. Una funzione $f:\mathbb{R}^2\to\mathbb{R}$ ha le seguenti proprietà: $f_x(0,0)=1$, $f_y(0,0)=-1$; $D_\mathbf{v}f(0,0)=2$ dove $\mathbf{v}=\left(\dfrac{1}{\sqrt{2}},\dfrac{1}{\sqrt{2}}\right)$. La funzione f è differenziabile in $(0,0)$?

8.* Dimostrare che se $f:A\to\mathbb{R}$, $A\subseteq\mathbb{R}^n$ aperto, ha derivate parziali limitate in A allora è continua in A.

9. Determinare per quale valore di α il piano tangente al grafico di $z=f(x,y)=$ $=\sin(\alpha x+y^2)$ nel punto $(0,\sqrt{\pi},0)$ è parallelo alla retta $x=y=2z$. Esistono valori di α per cui è perpendicolare?

10. Sia f differenziabile in $\mathbf{x}$ e sia $\mathbf{w}$ un vettore qualunque in $\mathbb{R}^n$; dimostrare che

$$\lim_{t\to0}\frac{1}{t}[f(\mathbf{x}+t\mathbf{w})-f(\mathbf{x})]=<\nabla f(\mathbf{x}),\mathbf{w}>\ .$$

11. Sia data la seguente funzione (vedi fig. 7.7)

$$f(x,y) = \begin{cases} 1 & \text{se } |y| > x^2 \text{ oppure se } y = 0 \\ 0 & \text{altrove} \end{cases}$$

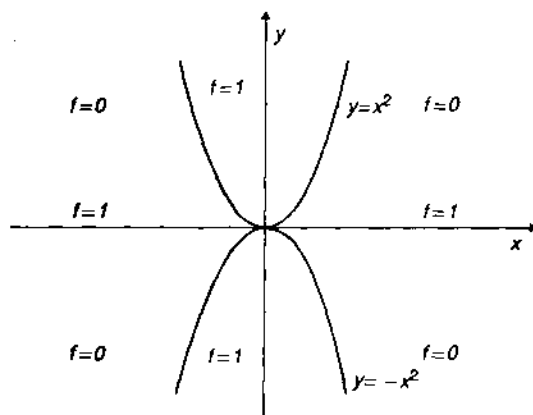



Fig. 7.7

Calcolare $D_{\mathbf{v}}f(0,0)$ per ogni versore $\mathbf{v} \in \mathbb{R}^2$; verificare se è valida la formula (1.3). La funzione f è differenziabile nell'origine?

12. Estendere il teorema di Schwarz alle derivate $D^2_{\mathbf{v}\mathbf{w}}f$ dove $\mathbf{v}$ e $\mathbf{w}$ sono versori qualunque di $\mathbb{R}^n$.

13.* Trovare tutte le funzioni $f : \mathbb{R}^2 \to \mathbb{R}$ differenziabili che soddisfano $\forall \mathbf{x} \in \mathbb{R}^2$ una delle seguenti condizioni:

i) $f_{x_2}(\mathbf{x}) = 0$

ii) $f_{x_1}(\mathbf{x}) + f_{x_2}(\mathbf{x}) = 0$

iii) $af_{x_1}(\mathbf{x}) + bf_{x_2}(\mathbf{x}) = 0 \quad (a, b \in \mathbb{R} \setminus \{0\})$.

Come possono essere interpretate fisicamente le soluzioni dei problemi *i), ii), iii)*?

[Suggerimento: *ii)* equivale a $D_{\mathbf{v}}f = 0$ se $\mathbf{v} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$; dunque f è costante lungo le rette $y - x = c$...; per *iii)* si ragiona come per *ii)* relativamente alla direzione $\mathbf{v} = \frac{1}{\sqrt{a^2+b^2}}(a, b)$].

14. Tra tutte le soluzioni dei problemi *i), ii), iii)* dell'esercizio precedente, trovare quella/e (se ne esistono) che soddisfano l'ulteriore condizione:

$$f(x_1, 0) = \sin x_1 \ .$$

15. Trovare le funzioni $f : \mathbb{R}^2 \to \mathbb{R}$, due volte differenziabili, che soddisfano la condizione

$$\frac{\partial^2 f}{\partial x \partial y} = 0 \quad \text{in } \mathbb{R}^2 .$$

Tra queste selezionare quella/e che soddisfa le ulteriori condizioni

$$f(0,y) = \sin y$$

$$f(x,0) = \cos x - 1 .$$

16. Scrivere le formule di Mac Laurin arrestate all'ordine $n = 3$ con il resto di Peano per le funzioni seguenti:

$a)$ $f(x,y) = \sin x \cdot \sin y$ $\qquad b)$ $f(x,y,z) = e^{x+y+z}$.

17. Vero o falso:

a) Sia $f : \mathbb{R}^2 \to \mathbb{R}$ tale che f_x, f_y siano definite in $\mathbb{R}^2$. Allora f è differenziabile in $(0,0)$.

b) Sia $f : \mathbb{R}^2 \to \mathbb{R}$ tale che: f_x ed f_y esistono in $\mathbb{R}^2$, f_y è continua in $\mathbb{R}^2$. Allora f è differenziabile in $\mathbb{R}^2$.

18. Sia $f : \mathbb{R}^n \to \mathbb{R}$ tale che $||\nabla f(\mathbf{x})|| \leq K$ per ogni $\mathbf{x} \in \mathbb{R}^n$.
Dimostrare che $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

$$|f(\mathbf{x}) - f(\mathbf{y})| \leq K||\mathbf{x} - \mathbf{y}|| .$$

È vero tale risultato per funzioni da $\mathbb{R}^n$ in $\mathbb{R}^m$, $m > 1$?

19. Sia $\mathbf{M} = (m_{jk})_{j,k=1,\dots,n}$ una matrice quadrata di ordine n.
Dimostrare che

$$\frac{\partial}{\partial m_{jk}} \det \mathbf{M} = M_{jk}$$

dove M_{jk} è il complemento algebrico dell'elemento m_{jk}.

20. Dimostrare che, se $f : C \to \mathbb{R}$, C cono aperto di $\mathbb{R}^n$, è differenziabile in C ed è positivamente omogenea di grado μ, allora le sue derivate parziali sono positivamente omogenee di grado $\mu - 1$.

21. Sia f convessa in $A \subseteq \mathbb{R}^n$ (A convesso). Dimostrare che, per ogni $c \in \mathbb{R}$, l'insieme (eventualmente vuoto),

$$E_c = \{\mathbf{x} \in A \ : \ f(\mathbf{x}) \leq c\}$$

è convesso.

22. Siano f, g convesse in A; $\alpha, \beta \in \mathbb{R}_+$. Dimostrare che $\alpha f + \beta g$ è convessa in A.

23. Sia f convessa in A, $F : \mathbb{R} \to \mathbb{R}$ convessa e *crescente*; dimostrare che $F \circ f$ è convessa in A.

24.* Diciamo che una forma quadratica $q(\mathbf{y})$ è semidefinita positiva se $q(\mathbf{y}) \geq 0$ per ogni $\mathbf{y} \in \mathbb{R}^n$.

Dimostrare che, se f è due volte differenziabile in A aperto convesso di $\mathbb{R}^n$, allora f è convessa in A se e solo se $d^2 f(\mathbf{x})$ è semidefinita positiva $\forall \mathbf{x} \in A$.

[Usare la formula di Taylor con il resto di Peano per il "solo se"].

2. FUNZIONI A VALORI VETTORIALI

2.1 Derivate e differenziali

Le nozioni di derivata e differenziale si estendono in modo naturale a funzioni con valori in $\mathbb{R}^m$, $m > 1$.

In generale penseremo i vettori scritti come vettori riga. D'altra parte se $\mathbf{x}$ è un vettore in $\mathbb{R}^n$ ed $\mathbf{M}$ una matrice $m \times n$, nel prodotto (righe per colonne) $\mathbf{M}\mathbf{x}$, $\mathbf{x}$ è pensato come vettore colonna. Anche se dovrebbe risultare chiaro dal contesto, preciseremo quando in una formula i vettori devono essere pensati come colonne.

Sia $\mathbf{f} : \mathbb{R}^n \supseteq A \to \mathbb{R}^m$, A aperto. Per ogni $\mathbf{x} \in A$, $\mathbf{f}(\mathbf{x})$ è un vettore $(f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x}))$ le cui componenti f_j, $j = 1, \ldots, m$ sono funzioni da A in $\mathbb{R}$.

Fissato un versore $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{f}$ è derivabile lungo la direzione $\mathbf{v}$ nel punto $\mathbf{x}$ se e solo se esistono $D_{\mathbf{v}} f_j(\mathbf{x})$ per ogni $j = 1, \ldots, m$ e

$$D_{\mathbf{v}}\mathbf{f} := (D_{\mathbf{v}} f_1, D_{\mathbf{v}} f_2, \ldots, D_{\mathbf{v}} f_m) .$$

Per quanto riguarda la differenziabilità, l'idea è sempre la stessa e cioè quella di approssimare l'incremento $\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x})$ (vettore di $\mathbb{R}^m$) con una trasformazione lineare in $\mathbf{h}$, da $\mathbb{R}^n$ in $\mathbb{R}^m$. Ora, ogni trasformazione di questo tipo è della forma $\mathbf{Mh}$ dove $\mathbf{M}$ è una matrice $m \times n$ ed il prodotto è quello "righe per colonne" (*). Siamo così condotti alla seguente definizione:

Definizione 2.1 - $\mathbf{f}$ *si dice differenziabile in* $\mathbf{x}$ *se esiste una matrice* $\mathbf{M}$ *di ordine* $m \times n$ *tale che* (vettori colonna):

$$\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x}) = \mathbf{Mh} + o(\|\mathbf{h}\|) \quad \text{per } \|\mathbf{h}\| \to 0 \tag{2.1}$$

per ogni $\mathbf{h} \in \mathbb{R}^n$ *con* $\mathbf{x}+\mathbf{h} \in A$. *L'applicazione lineare da* $\mathbb{R}^n$ *in* $\mathbb{R}^m$ *data da:*

$$\mathbf{h} \mapsto \mathbf{Mh}$$

prende il nome di **differenziale** *di* f *in* $\mathbf{x}$ *e viene indicata col simbolo* $d\mathbf{f}(\mathbf{x})$.

Il teorema 1.1 continua a valere; per quanto riguarda "l'identificazione" della matrice $\mathbf{M}$ osserviamo che scrivendo la (2.1) per la componente f_k di $\mathbf{f}$ si ottiene

$$f_k(\mathbf{x}+\mathbf{h}) - f_k(\mathbf{x}) = m_{k1}h_1 + m_{k2}h_2 + \ldots + m_{kn}h_n + o(\|\mathbf{h}\|) . \tag{2.2}$$

(*) $\mathbb{R}^n$ ed $\mathbb{R}^m$ si intendono riferiti alla base canonica.

La (2.2) indica che f_k è differenziabile e che la k-esima riga di $\mathbf{M}$ è $\nabla f_k(\mathbf{x})$, in base alla (1.3). Dunque si ha:

$$\mathbf{M} = \begin{pmatrix} \nabla f_1(\mathbf{x}) \\ \nabla f_2(\mathbf{x}) \\ \vdots \\ \nabla f_m(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} D_{x_1} f_1(\mathbf{x}) & D_{x_2} f_1(\mathbf{x}) & \dots & D_{x_n} f_1(\mathbf{x}) \\ D_{x_1} f_2(\mathbf{x}) & D_{x_2} f_2(\mathbf{x}) & \dots & D_{x_n} f_2(\mathbf{x}) \\ \vdots & & & \\ D_{x_1} f_m(\mathbf{x}) & D_{x_2} f_m(\mathbf{x}) & \dots & D_{x_n} f_m(\mathbf{x}) \end{pmatrix} . \quad (2.3)$$

La matrice (2.3) prende il nome di *matrice Jacobiana* (*) di $\mathbf{f}$ in $\mathbf{x}$ e si indica con i simboli

$$\mathbf{J}_f(\mathbf{x}) \ , \ \mathbf{Df}(\mathbf{x}) \ , \ \mathbf{f}'(\mathbf{x}) \ .$$

Scrivendo la (2.1) per f_k, $k = 1, \dots, m$, si vede che $\mathbf{f}$ *è differenziabile in* $\mathbf{x}$ *se e solo se lo è ognuna delle sue componenti*. Continua dunque a valere il teorema 1.2.

Ragionando sulle componenti è poi possibile definire derivate e differenziali di ordine superiore e verificare la validità dei teoremi 1.3 e 1.4. Per quanto riguarda la formula di Taylor, per $\mathbf{f}$ vettore continua a valere il teorema 1.8; non può invece essere valido il teorema 1.5 (e di conseguenza anche il teorema del valor medio) come mostra il seguente semplice esempio.

***Esempio* 2.1** - Sia $\mathbf{f}(x) = (\cos x, \sin x)$, $x \in [0, 2\pi]$. Allora $\mathbf{f}(2\pi) - \mathbf{f}(0) = \mathbf{0}$ ma non esiste alcun $\overline{x} \in [0, 2\pi]$ per cui $\mathbf{f}'(\overline{x}) = (-\sin \overline{x}, \cos \overline{x})$ sia nullo.

Concludiamo il paragrafo con una serie di esempi di funzioni a valori vettoriali evidenziandone il significato geometrico e fisico.

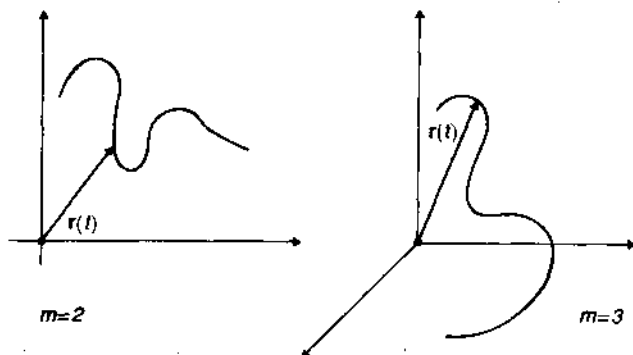



Fig. 7.8

***Esempio* 2.2** - *Funzioni vettoriali di variabile reale.*

Sia $I \subseteq \mathbb{R}$ un intervallo. Per funzioni da I in $\mathbb{R}^m$ conviene usare t come

(*) Dal nome del matematico Carl Gustave Jacobi.

variabile indipendente e $\mathbf{r}(t) = (x_1(t), x_2(t), \ldots, x_m(t))$ come notazione per la funzione e le sue componenti. Nei casi $n = 2$ e $n = 3$ useremo la notazione più comune $\mathbf{r}(t) = (x(t), y(t))$ e $\mathbf{r}(t) = (x(t), y(t), z(t))$, rispettivamente.

In questi casi $\mathbf{r}(t)$ è interpretabile come posizione di un punto al tempo t nel piano o nello spazio rispettivamente; al variare di t in I, $\mathbf{r}(t)$ descrive una "linea" (fig. 7.8).

Ad esempio:

$$\mathbf{r}(t) = (x_0 + \alpha t, y_0 + \beta t, z_0 + \gamma t) \ , \ t \in \mathbb{R}, \ \alpha^2 + \beta^2 + \gamma^2 = 1 \ ,$$

al variare di t in $\mathbb{R}$ descrive una retta passante per (x_0, y_0, z_0) con coseni direttori α, β, γ.

$\mathbf{r}(t) = (x_0 + R\cos t, y_0 + R\sin t)$, $t \in [0, 2\pi], R > 0$, descrive una circonferenza con centro (x_0, y_0) e raggio R (*).

$\mathbf{r}(t) = (x_0 + R\cos t, y_0 + R\sin t, kt)$, $t \in \mathbb{R}$, $R > 0, k \in \mathbb{R}$, descrive un'elica cilindrica (fig. 7.9).

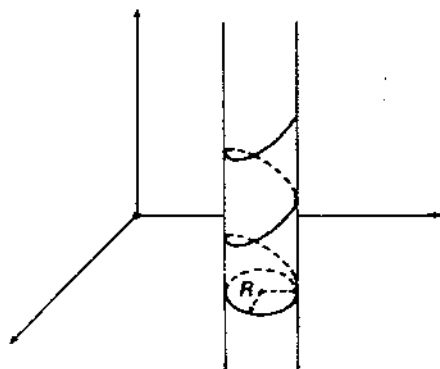



Fig. 7.9

$\mathbf{r}(t)$ è differenziabile se tutte le sue componenti $x_1, \ldots, x_n$ sono derivabili. La matrice Jacobiana di $\mathbf{r}$ è il vettore

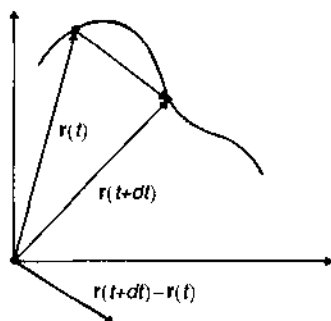
$$\mathbf{r}'(t) = (x_1'(t), x_2'(t), \ldots, x_n'(t)) \ .$$

Cinematicamente $\mathbf{r}'(t)$ rappresenta la velocità del punto ($\|\mathbf{r}'(t)\|$ è la velocità scalare). Così $\mathbf{r}''(t)$ sarà la sua accelerazione.

Geometricamente $\mathbf{r}'(t)$ è un vettore *tangente* alla curva nel punto $\mathbf{r}(t)$ come è illustrato in figura 7.10 nel caso $m = 2$.

Se $\mathbf{r}, \mathbf{u} : I \to \mathbb{R}^m$ sono differenziabili in I risulta differenziabile anche la funzione $f(t) = <\mathbf{r}(t), \mathbf{u}(t)>$ da I in $\mathbb{R}$, prodotto scalare di $\mathbf{r}$ ed $\mathbf{u}$; infatti

(*) L'allievo consideri il caso $t \in [0, 4\pi]$ o più in generale $t \in [0, 2k\pi]$, $k \in \mathbb{N}_+$.

Fig. 7.10 Per $dt \to 0$ il vettore $\dfrac{\mathbf{r}(t+dt)-\mathbf{r}(t)}{dt}$ tende a una posizione limite, tangente alla linea in $\mathbf{r}(t)$.

$$f(t) = \sum_{j=1}^{m} x_j(t) u_j(t) \quad (\mathbf{u}(t) = (u_1(t) \ldots u_m(t))\,.$$

Vale la formula, come facilmente si verifica

$$f'(t) = <\mathbf{r}'(t), \mathbf{u}(t)> + <\mathbf{r}(t), \mathbf{u}'(t)>$$

che generalizza la formula per la derivata del prodotto di due funzioni reali.
Torneremo estesamente sulle funzioni da I in $\mathbb{R}^3$ nel secondo volume.

***Esempio* 2.3 - *Funzioni da* $\mathbb{R}^2$ *in* $\mathbb{R}^3$.**

Sia $A \subseteq \mathbb{R}^2$, A aperto. Per le funzioni da A in $\mathbb{R}^3$ usiamo la notazione $\mathbf{r}(u,v) = (x(u,v), y(u,v), z(u,v))$.

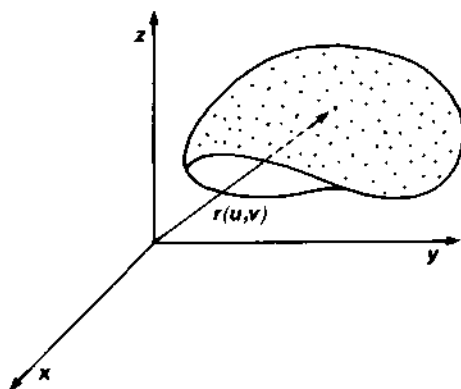



Fig. 7.11

Al variare di (u,v) in A, $\mathbf{r}(u,v)$ descrive una "superficie" (*) nello spazio (fig. 7.11).

(*) I concetti di linea e di superficie verranno trattati rigorosamente in seguito.

Se $\mathbf{r}$ è differenziabile in A la sua matrice Jacobiana è:

$$\mathbf{r}' = \begin{pmatrix} x_u & x_v \\ y_u & y_v \\ z_u & z_v \end{pmatrix}$$

I vettori $\mathbf{r}_u = (x_u, y_u, z_u)$ ed $\mathbf{r}_v = (x_v, y_v, z_v)$ sono vettori tangenti alle linee coordinate $\mathbf{r}(u, \overline{v})$ e $\mathbf{r}(\overline{u}, v)$ sulla superficie, corrispondenti a $\overline{v}$ =costante e $\overline{u}$ =costante nel piano (u, v) (fig. 7.12).

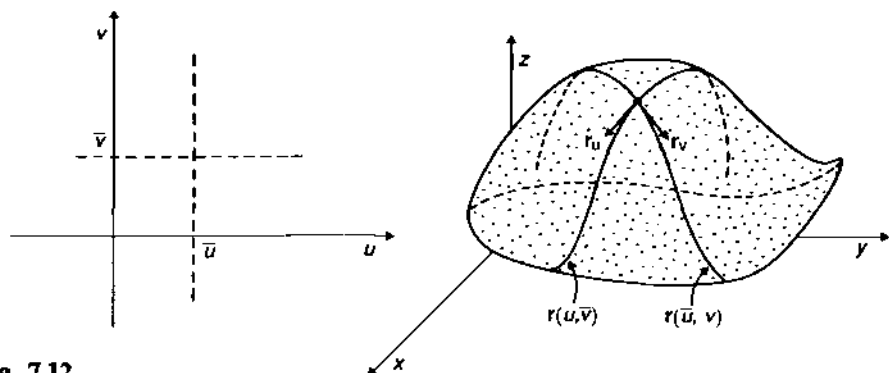



Fig. 7.12

Un ruolo importante nella teoria delle superfici è svolto dal prodotto vettoriale $\mathbf{r}_u \wedge \mathbf{r}_v$, vettore normale alla superficie. Anche su queste questioni ritorneremo ampiamente in seguito.

Esempio 2.4 - Funzioni da $\mathbb{R}^n$ *in* $\mathbb{R}^n$.

Se si scrive $\mathbf{f} = (x_1(u_1, \ldots, u_n), x_2(u_1, \ldots, u_n), \ldots, x_n(u_1, \ldots, u_n))$ $\mathbf{f}$ è interpretabile come un cambiamento di variabili in $\mathbb{R}^n$; in questo caso si preferisce scrivere

$$\begin{cases} x_1 = x_1(u_1, u_2, \ldots, u_n) \\ x_2 = x_2(u_1, u_2, \ldots, u_n) \\ \vdots \\ x_n = x_n(u_1, u_2, \ldots, u_n) \end{cases} \tag{2.4}$$

e si usa per $\mathbf{f}$ il termine *trasformazione* (di coordinate) in $\mathbb{R}^n$.

Un esempio è costituito dai cambiamenti di variabili lineari, esprimibili nella forma

$$\mathbf{x} = \mathbf{M}\mathbf{u} \text{ (vettori colonna) oppure } \mathbf{x} = \mathbf{u}\mathbf{M}^T \text{ (vettori riga)} \tag{2.5}$$

dove $\mathbf{M}$ è una matrice quadrata di ordine n; $\mathbf{M}^T$ indica la matrice trasposta di $\mathbf{M}$, cioè la matrice che si ottiene da $\mathbf{M}$ scambiando le righe con le colonne.

In questo caso la matrice Jacobiana della trasformazione coincide con $\mathbf{M}$, infatti se m_{jk} sono gli elementi di $\mathbf{M}$, la (2.5) implica che

$$x_j = \sum_{k=1}^{n} m_{jk} u_k$$

e quindi $\partial x_j / \partial u_k = m_{jk}$.

Si noti che la (2.4) è $1 \leftrightarrow 1$ da $\mathbb{R}^n$ in $\mathbb{R}^n$ se e solo se $\mathbf{M}$ è non singolare (cioè $\det \mathbf{M} \neq 0$) ovvero se e solo se la matrice Jacobiana della trasformazione è non singolare.

Anche per una trasformazione generale come la (2.4) è importante, come vedremo, il determinante della matrice Jacobiana (o *determinante Jacobiano*); per esso è preferibile usare la notazione

$$\frac{\partial(x_1, x_2, \ldots, x_n)}{\partial(u_1, u_2, \ldots, u_n)} .$$

Esempi notevoli di trasformazioni sono i seguenti.

***Esempio* 2.5** - Trasformazioni lineari particolarmente importanti sono quelle rappresentate da una matrice $\mathbf{M}$ *ortogonale*, cioè tale che $\mathbf{M}^{-1} = \mathbf{M}^T$.

In altri termini $\mathbf{M} \cdot \mathbf{M}^T = \mathbf{M}^T \cdot \mathbf{M} = \mathbf{I}_n$ dove $\mathbf{I}_n$ è la matrice identità di ordine n:

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix} .$$

Esempi di matrici ortogonali nel piano sono

$$\mathbf{M}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ e } \mathbf{M}_2 = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} .$$

La prima rappresenta una riflessione rispetto alla bisettrice $x_1 = x_2$; infatti

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix} ;$$

la seconda una rotazione antioraria di angolo θ e centro nell'origine (fig. 7.13).

Notevoli proprietà delle matrici ortogonali sono le seguenti (lasciamo la verifica al lettore):

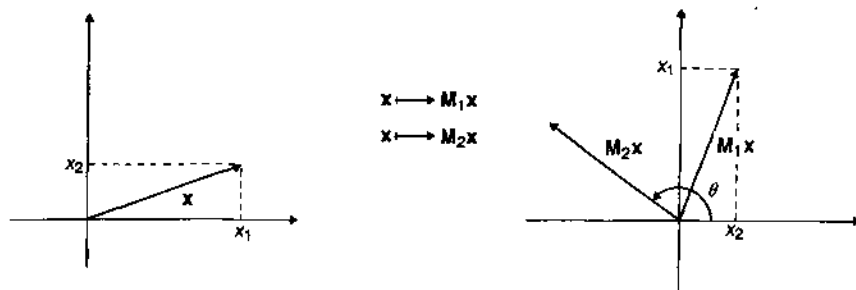



Fig. 7.13

i) $\det \mathbf{M} = \pm 1$;

ii) l'insieme dei vettori riga (colonna) costituisce una base ortogonale per $\mathbb{R}^n$;

iii) $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \quad , \quad < \mathbf{M}\mathbf{x}, \mathbf{M}\mathbf{y} > = < \mathbf{x}, \mathbf{y} >$

ed in particolare $\|\mathbf{M}\mathbf{x}\| = \|\mathbf{x}\|$ (la lunghezza di un vettore non varia per trasformazioni ortogonali).

Esempio 2.6 - *Coordinate polari nel piano rispetto al polo* (x_0, y_0)

$$\begin{cases} x = x_0 + \rho\cos\theta \\ y = y_0 \ + \rho\sin\theta \end{cases} \qquad (\rho, \theta) \in (0 + \infty) \times [0, 2\pi) \ .$$

Si ha

$$\frac{\partial(x,y)}{\partial(\rho,\theta)} = \begin{vmatrix} x_\rho & x_\theta \\ y_\rho & y_\theta \end{vmatrix} = \begin{vmatrix} \cos\theta & -\rho\sin\theta \\ \sin\theta & \rho\cos\theta \end{vmatrix} = \rho \ .$$

Esempio 2.7 - *Coordinate cilindriche nello spazio rispetto all'asse* $x = x_0$, $y = y_0$.

$$\begin{cases} x = x_0 + \rho\cos\theta \\ y = y_0 + \rho\sin\theta \\ z = t \end{cases} \qquad (\rho, \theta, t) \in (0, +\infty) \times [0, 2\pi) \times \mathbb{R} \ .$$

Si ha:

$$\frac{\partial(x,y,z)}{\partial(\rho,\theta,t)} = \begin{vmatrix} x_\rho & x_\theta & x_t \\ y_\rho & y_\theta & y_t \\ z_\rho & z_\theta & z_t \end{vmatrix} = \begin{vmatrix} \cos\theta & -\rho\sin\theta & 0 \\ \sin\theta & \rho\cos\theta & 0 \\ 0 & 0 & 1 \end{vmatrix} = \rho \ .$$

Esempio 2.8 - *Coordinate polari o sferiche nello spazio rispetto al polo* $(0, 0, 0)$.

$(\rho, \psi, \theta) \in (0, +\infty) \times [0, \pi] \times [0, 2\pi)$.

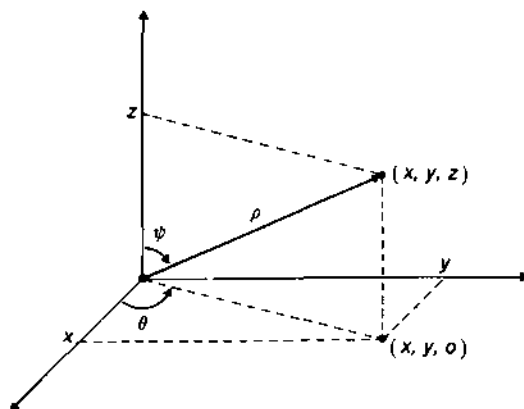



Fig. 7.14

$$\begin{cases} x = \rho \sin \psi \cos \theta \\ y = \rho \sin \psi \sin \theta \\ z = \rho \cos \psi \end{cases}$$

Si ha

$$\frac{\partial(x,y,z)}{\partial(\rho,\psi,\theta)} = \begin{vmatrix} \sin\psi\cos\theta & \rho\cos\psi\cos\theta & -\rho\sin\psi\sin\theta \\ \sin\psi\sin\theta & \rho\cos\psi\sin\theta & \rho\sin\psi\cos\theta \\ \cos\psi & -\rho\sin\psi & 0 \end{vmatrix} = \rho^2 \sin\psi \ .$$

***Esempio* 2.9** - Ancora funzioni da $\mathbb{R}^n$ ad $\mathbb{R}^n$.

Oltre che come trasformazioni di variabili le funzioni $\mathbf{f} : \mathbb{R}^n \supseteq A \to \mathbb{R}^n$ si possono interpretare come *campi di vettori*.

Un esempio per $n = 2$ è il campo di velocità per il moto piano stazionario di un fluido:

$$\mathbf{v}(x,y) = (v_1(x,y), v_2(x,y))$$

che assegna la velocità della particella di fluido che si trova nel punto di coordinate (x,y).

Un altro esempio è il campo elettrostatico

$$\mathbf{E}(x,y,z) = (E_1(x,y,z), E_2(x,y,z), E_3(x,y,z)) \ .$$

Per i campi vettoriali derivabili è definito un operatore differenziale che si chiama *divergenza* e si indica con "div" e che in coordinate cartesiane è assegnato dalla formula

$$\operatorname{div} \mathbf{f} := \frac{\partial f_1}{\partial x_1} + \frac{\partial f_2}{\partial x_2} + \ldots + \frac{\partial f_n}{\partial x_n} \ .$$

Ai campi vettoriali $\mathbf{f}$ derivabili in $\mathbb{R}^3$ si può associare un altro vettore, il *rotore* di $\mathbf{f}$, denotato con $\operatorname{rot}\mathbf{f}$ oppure $\operatorname{curl}\mathbf{f}$, le cui componenti in coordinate cartesiane sono le seguenti

$$\operatorname{rot}\mathbf{f} := \left(\frac{\partial f_3}{\partial x_2} - \frac{\partial f_2}{\partial x_3}, \frac{\partial f_1}{\partial x_3} - \frac{\partial f_3}{\partial x_1}, \frac{\partial f_2}{\partial x_1} - \frac{\partial f_1}{\partial x_2}\right) .$$

I campi vettoriali a divergenza nulla sono detti *solenoidali*; quelli a rotore nullo, *irrotazionali*.

L'importanza di divergenza e rotore sarà più chiara in seguito.

2.2 Differenziale delle funzioni composte

Vediamo ora come si comporta il differenziale rispetto all'operazione di composizione.

Vale il seguente importante risultato.

■ **Teorema 2.1** - *Siano $A \subseteq \mathbb{R}^n$, $B \subseteq \mathbb{R}^m$ aperti e*

$$\mathbf{f} : A \to B \ , \ \mathbf{g} : B \to \mathbb{R}^p \ .$$

Se $\mathbf{f}$ è differenziabile in $\mathbf{x} \in A$ e $\mathbf{g}$ è differenziabile in $\mathbf{f}(\mathbf{x})$, allora la funzione composta $\mathbf{g} \circ \mathbf{f} : A \to \mathbb{R}^p$ è differenziabile in $\mathbf{x}$ e vale la formula:

$$\mathbf{J}_{\mathbf{g}\circ\mathbf{f}}(\mathbf{x}) = \mathbf{J}_{\mathbf{g}}(\mathbf{f}(\mathbf{x})) \cdot \mathbf{J}_{\mathbf{f}}(\mathbf{x}) \ . \tag{2.6}$$

Notiamo che $\mathbf{J}_{\mathbf{g}\circ\mathbf{f}}$ è di ordine $p \times n$, $\mathbf{J}_{\mathbf{g}}$ di ordine $p \times m$ e $\mathbf{J}_{\mathbf{f}}$ di ordine $m \times n$.

La dimostrazione del teorema ricalca quella del caso $n = m = p = 1$ per cui la omettiamo. Ricaviamo invece dalla (2.6) una formula esplicita per le derivate di ogni componente di $\mathbf{g} \circ \mathbf{f}$.

A tale scopo poniamo

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) \ , \ \varphi(\mathbf{x}) = \mathbf{g} \circ \mathbf{f}(\mathbf{x}) = \mathbf{g}(\mathbf{y}) \ .$$

Dalla (2.6) si ricava:

$$\frac{\partial \varphi_k}{\partial x_j} = \sum_{i=1}^{m} \frac{\partial g_k}{\partial y_i} \cdot \frac{\partial y_i}{\partial x_j} \qquad \begin{array}{l} k = 1, \ldots, p \\ j = 1, \ldots, n \end{array} . \tag{2.7}$$

Usando le altre notazioni per la matrice Jacobiana si può scrivere la (2.6) in una delle forme seguenti

$$\mathbf{D}\varphi(\mathbf{x}) = \mathbf{D}\mathbf{g}(\mathbf{y}) \cdot \mathbf{D}\mathbf{f}(\mathbf{x})$$

oppure

$$\varphi'(\mathbf{x}) = \mathbf{g}'(\mathbf{y}) \cdot \mathbf{f}'(\mathbf{x}) \ .$$

***Esempio* 2.10** - Sia $f : \mathbb{R}^2 \setminus \{0,0\} \to \mathbb{R}$ differenziabile; se $f = f(x,y)$ poniamo $\tilde{f}(\rho,\theta) = \tilde{f}(\rho\cos\theta, \rho\sin\theta)$, effettuando così un passaggio a coordinate polari.

Vogliamo esprimere le derivate di $\tilde{f}$ rispetto a ρ e θ mediante quelle di f rispetto ad x ed y.

Dalla (2.7) abbiamo:

$$\frac{\partial \tilde{f}}{\partial \rho} = \frac{\partial f}{\partial x} \cdot \frac{\partial x}{\partial \rho} + \frac{\partial f}{\partial y} \cdot \frac{\partial y}{\partial \rho} = \frac{\partial f}{\partial x}\cos\theta + \frac{\partial f}{\partial y}\sin\theta$$

$$\frac{\partial \tilde{f}}{\partial \theta} = \frac{\partial f}{\partial x} \cdot \frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y} \cdot \frac{\partial y}{\partial \theta} = \frac{\partial f}{\partial x}(-\rho\sin\theta) + \frac{\partial f}{\partial y}(\rho\cos\theta) .$$

Le stesse formule si trovano applicando direttamente la (2.6) tenendo presente che

$$\begin{pmatrix} \cos\theta & -\rho\sin\theta \\ \sin\theta & \rho\cos\theta \end{pmatrix}$$

è la matrice Jacobiana della trasformazione in polari:

$$\left(\frac{\partial \tilde{f}}{\partial \rho} \ , \ \frac{\partial \tilde{f}}{\partial \theta}\right) = \left(\frac{\partial f}{\partial x} \ , \ \frac{\partial f}{\partial y}\right) \cdot \begin{pmatrix} \cos\theta & -\rho\sin\theta \\ \sin\theta & \rho\cos\theta \end{pmatrix} .$$

***Esempio* 2.11** - Sia $f : \mathbb{R}^n \to \mathbb{R}$ due volte differenziabile. Operiamo il cambio di variabile $\mathbf{x} = \mathbf{M}\mathbf{y}$ (vettori colonna) dove $\mathbf{M}$ è una matrice quadrata di ordine n.

Poniamo

$$\tilde{f}(\mathbf{y}) = f(\mathbf{M}\mathbf{y}) .$$

Vogliamo stabilire quale relazione sussiste tra le matrici Hessiane di f ed $\tilde{f}$.

Usando la formula (2.7) abbiamo (ricordando l'esempio 2.4):

$$\frac{\partial \tilde{f}}{\partial y_j}(\mathbf{y}) = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(\mathbf{x}) \cdot \frac{\partial x_k}{\partial y_j} = \sum_{k=1}^{n} \frac{\partial f}{\partial x_k}(\mathbf{x}) m_{kj} .$$

Riapplicando la stessa formula per $\dfrac{\partial \tilde{f}}{\partial y_j}$:

$$\frac{\partial^2 \tilde{f}}{\partial y_i \partial y_j}(\mathbf{y}) = \sum_{s=1}^{n} \sum_{k=1}^{n} \frac{\partial^2 f}{\partial x_s \partial x_k}(\mathbf{x}) m_{si} m_{kj} . \tag{2.8}$$

Se $\mathbf{M}^T$ indica la trasposta di $\mathbf{M}$, la (2.8) si riscrive nella forma seguente:

$$\mathbf{H}_{\tilde{f}}(\mathbf{y}) = \mathbf{M}^T \cdot \mathbf{H}_f(\mathbf{x}) \cdot \mathbf{M} \ , \tag{2.9}$$

che è la relazione cercata.

2.3 Funzioni da $\mathbb{C}$ in $\mathbb{C}$

In questo paragrafo ci occuperemo di funzioni complesse di variabile complessa. Sia dunque $w = f(z)$ una funzione definita in un aperto $A \subseteq \mathbb{C}$ a valori in $\mathbb{C}$.

Introducendo parte reale e parte immaginaria di z ed f possiamo identificare $\mathbb{C}$ con $\mathbb{R}^2$ e considerare f come una funzione da $\mathbb{R}^2$ in $\mathbb{R}^2$. In altri termini, posto $z = x + iy, u(x,y) = \operatorname{Re} f(z), v(x,y) = \operatorname{Im} f(z)$ avremo $f(z) = \mathbf{F}(x,y) = (u(x,y), v(x,y))$. Ad esempio se $f(z) = z^2, u(x,y) = x^2 - y^2$ e $v(x,y) = 2xy$.

Per $\mathbf{F}$ sono dunque definite tutte le nozioni introdotte nei paragrafi precedenti ed in particolare la differenziabilità.

D'altra parte l'identificazione tra $\mathbb{C}$ ed $\mathbb{R}^2$ è possibile solo a livello di spazio vettoriale metrico (su $\mathbb{R}$) mentre sappiamo che $\mathbb{C}$ è dotato altresì della struttura di campo, che permette moltiplicazioni e divisioni tra numeri complessi.

È quindi ragionevole pensare di estendere la nozione di derivata, così come è stata introdotta per le funzioni reali di *una* variabile reale, alle funzioni complesse di *una* variabile complessa.

Definizione 2.2 - *Sia $f : \mathbb{C} \supseteq A \to \mathbb{C}$, A aperto e $z_0 \in A$. Se esiste finito il limite:*

$$\lim_{z \to z_0} \frac{f(z) - f(z_0)}{z - z_0} \qquad (z \in A)$$

questo viene detto derivata complessa di f in z_0 ed indicato con uno dei simboli $f'(z_0)$, $Df(z_0)$. La funzione f si dice derivabile in senso complesso o dotata di derivata in senso complesso in z_0.

Se f ha derivata complessa in ogni punto di A, f si dice **olomorfa** in A.

***Esempio* 2.12** - Sia $f(z) = z^n$. Allora

$$\frac{f(z) - f(z_0)}{z - z_0} = z^{n-1} + z^{n-2} z_0 + \cdots + z z_0^{n-2} + z_0^{n-1} \longrightarrow n z_0^{n-1}$$

per $z \to z_0$ e quindi $f'(z_0) = n z_0^{n-1}$ esattamente come nel caso reale.

La funzione $f(z) = z^n$ è olomorfa in $\mathbb{C}$.

Le regole di derivazione per somma, prodotto, quoziente e composizione si estendono senza difficoltà alla derivata complessa.

Ciò che ci proponiamo è di esaminare la relazione tra la derivabilità in senso complesso di f e la differenziabilità della corrispondente $\mathbf{F}$.

Sia $z_0 = x_0 + iy_0$. La differenziabilità di $\mathbf{F}$ in (x_0, y_0) equivale ad affermare che

$$u(x,y) = u(x_0,y_0) + u_x(x_0,y_0)(x-x_0) + u_y(x_0,y_0)(y-y_0) + \\ + o\left(\sqrt{(x-x_0)^2+(y-y_0)^2}\right)$$

e

$$v(x,y) = v(x_0,y_0) + v_x(x_0,y_0)(x-x_0) + v_y(x_0,y_0)(y-y_0) + \\ + o\left(\sqrt{(x-x_0)^2+(y-y_0)^2}\right) .$$

Possiamo scrivere le due equazioni precedenti in forma compatta moltiplicando la seconda per i, sommando membro a membro ed usando la notazione complessa; si ricava:

$$f(z) = f(z_0) + f_x(z_0)(x-x_0) + f_y(z_0)(y-y_0) + o(|z-z_0|) , \tag{2.10}$$

per $z \to z_0$.

La derivabilità complessa di f in z_0 equivale invece a scrivere

$$f(z) = f(z_0) + f'(z_0)(z-z_0) + o(|z-z_0|), \text{ per } z \to z_0 . \tag{2.11}$$

Il seguente teorema indica che la (2.11) implica la (2.10) insieme ad una relazione quantitativa su f_x e f_y:

■ **Teorema 2.2** - *Siano $f : \mathbb{C} \supseteq A \to \mathbb{C}, z_0 \in A$; f è derivabile in senso complesso in z_0 se e solo se $\mathbf{F}$ è differenziabile in (x_0, y_0) e vale la relazione (di Cauchy–Riemann):*

$$f_y(z_0) = if_x(z_0) . \tag{2.12}$$

Dimostrazione - Sia f derivabile in senso complesso. Dalla (2.11) si ha :

$$f(z) = f(z_0) + f'(z_0)(x-x_0) + if'(z_0)(y-y_0) + o(|z-z_0|)$$

che implica la (2.10) con $f'(z_0) = f_x(z_0)$ e $if'(z_0) = f_y(z_0)$. Vale dunque anche la (2.12).

Sia $\mathbf{F}$ differenziabile in (x_0, y_0); vale allora la (2.10), che, usando la relazione di Cauchy–Riemann, si può scrivere nella forma seguente:

$$f(z) = f(z_0) + f_x(z_0)(z-z_0) + o(|z-z_0|) .$$

Quest'ultima relazione indica che f è derivabile con $f'(z_0) = f_x(z_0)$. □

Osserviamo che, nelle ipotesi del teorema, si ottiene $f'(z_0) = f_x(z_0) = -if_y(z_0)$.

Esempi

2.13. Sia $f(z) = e^z = e^x e^{iy} = e^x(\cos y + i \sin y)$. Le funzioni $u(x,y) = e^x \cos y, v(x,y) = e^x \sin y$ sono di classe $C^\infty(\mathbb{R}^2)$, dunque differenziabili in tutto il piano. Inoltre:

$$f_x(z) = e^x(\cos y + i \sin y)$$

$$f_y(z) = e^x(-\sin y + i \cos y) = i f_x(z) .$$

Quindi la (2.12) è verificata e $f(z)$ è olomorfa in $\mathbb{C}$; inoltre $f'(z) = f_x(z) = e^z$.

2.14. Sia $f(z) = \sin z$; poiché $\sin z = \dfrac{e^{iz} - e^{-iz}}{2i}$ si deduce che $\sin z$ è olomorfa in $\mathbb{C}$ e $f'(z) = \dfrac{e^{iz} + e^{-iz}}{2} = \cos z$.

Analogamente se $f(z) = \cos z, f$ è olomorfa in $\mathbb{C}$ e $f'(z) = -\sin z$.

2.15. Sia $f(z) = \overline{z} = x - iy$. Le funzioni $u(x,y) = x$ e $v(x,y) = -y$ sono di classe $C^\infty(\mathbb{R}^2)$. Inoltre:

$$f_x(z) = 1 \ , \ f_y(z) = -i .$$

Perciò $\overline{z}$ non è derivabile in alcun punto di $\mathbb{C}$.

La condizione di Cauchy–Riemann equivale alle due seguenti, ottenute uguagliando le parti reali e quelle immaginarie nella (2.12):

$$\begin{cases} u_x(x,y) = v_y(x,y) \\ u_y(x,y) = -v_x(x,y) . \end{cases} \tag{2.13}$$

Abbiamo visto che la derivabilità in senso complesso implica la differenziabilità di u e v e le equazioni (2.13).

In realtà implica molto di più ed infatti si dimostra che una funzione olomorfa in un aperto A del piano complesso ha derivate (in senso complesso) di *ogni* ordine; ciò implica che u, v siano di classe $C^\infty(A)$.

Deriviamo allora la prima delle (2.13) rispetto ad x e la seconda rispetto ad y; si ottiene:

$$\begin{cases} u_{xx}(x,y) = v_{xy}(x,y) \\ u_{yy}(x,y) = -v_{yx}(x,y) \end{cases} \qquad (x,y) \in A . \tag{2.14}$$

Per il teorema di Schwarz, $v_{xy}(x,y) = v_{yx}(x,y)$ e quindi, sommando le (2.14) membro a membro, si ricava l'importante risultato

$$\Delta u = u_{xx} + u_{yy} = 0$$

e cioè che u è *armonica* in A. Analogamente, deriviamo la prima delle (2.13) rispetto ad y, la seconda rispetto ad x e sottraiamo membro a membro; si ottiene

$$\Delta v = v_{xx} + v_{yy} = 0$$

ovvero, anche v è *armonica* in A. Se $f = u + iv$ è olomorfa in A, u e v si dicono *armoniche coniugate*.

Ad esempio $x^2 - y^2$ e $2xy$ sono armoniche coniugate; così pure $e^x \cos y$ ed $e^x \sin y$.

La teoria delle funzioni olomorfe è ricca di una grande quantità di importanti risultati, che non possono essere contenuti nei limiti della presente trattazione.

2.4 Il teorema di inversione locale

Se $\mathbf{l}$ è una funzione (trasformazione) lineare da $\mathbb{R}^n$ in $\mathbb{R}^n$, fissato il sistema di riferimento in $\mathbb{R}^n$, essa è rappresentata da una matrice $\mathbf{M}$, quadrata di ordine n. Ovvero se $\mathbf{x} \in \mathbb{R}^n$,

$$\mathbf{l}(\mathbf{x}) = \mathbf{M}\mathbf{x} \qquad \text{(vettori colonna)} .$$

Un noto teorema di algebra lineare (*Teorema di Cramer*) afferma che $\mathbf{l}$ realizza una corrispondenza biunivoca tra $\mathbb{R}^n$ ed $\mathbb{R}^n$ se e solo se $\mathbf{M}$ è non singolare; in tal caso per $\forall\, \mathbf{y} \in \mathbb{R}^n$ l'equazione (sistema di n equazioni lineari)

$$\mathbf{M}\mathbf{x} = \mathbf{y}$$

ha esattamente una soluzione $\mathbf{x} = \mathbf{M}^{-1}\mathbf{y}$.

Ciò equivale ad affermare che la funzione inversa di $\mathbf{l}, \mathbf{l}^{-1}$, pure lineare, è rappresentata da $\mathbf{M}^{-1}$. Di conseguenza la matrice Jacobiana di $\mathbf{l}^{-1}, \mathbf{M}^{-1}$, coincide con l'inversa della matrice Jacobiana di $\mathbf{l}$.

Il problema dell'inversione delle trasformazioni lineari da $\mathbb{R}^n$ in $\mathbb{R}^n$ è quindi completamente risolto.

Ci proponiamo di affrontare lo stesso problema nel caso *non lineare*; sia $\mathbf{f} : \mathbb{R}^n \supseteq A \to \mathbb{R}^n$, A aperto; sia $B = \mathrm{im}(\mathbf{f})$ (immagine di $\mathbf{f}$); vogliamo studiare l'invertibilità di $\mathbf{f}$ o, in altre parole, risolvere l'equazione

$$\mathbf{f}(\mathbf{x}) = \mathbf{y} \qquad \text{per } \mathbf{y} \in B . \tag{2.15}$$

Scriviamo per una volta la (2.15) esplicitamente:

$$\begin{cases} f_1(x_1, x_2, \dots, x_n) = y_1 \\ f_2(x_1, x_2, \dots, x_n) = y_2 \\ \vdots \\ f_n(x_1, x_2, \dots, x_n) = y_n . \end{cases} \tag{2.15'}$$

Mentre nel caso lineare il teorema di Cramer è *globale* ($\mathbf{l} : \mathbb{R}^n \xleftrightarrow{1-1} \mathbb{R}^n$), nel caso non lineare il principale risultato è di tipo *locale*. La dicotomia *globale–locale* associata a *lineare–non lineare* si presenta spesso in matematica. Risultato di tipo locale, nel caso specifico della (2.15), significa fissare un punto $\mathbf{x}_0 \in A$, considerare l'immagine $\mathbf{y}_0 = \mathbf{f}(\mathbf{x}_0)$ e cercare un intorno V di $\mathbf{x}_0$ ed uno W di $\mathbf{y}_0$ che siano in corrispondenza biunivoca tramite $\mathbf{f}$ (Fig. 7.15).

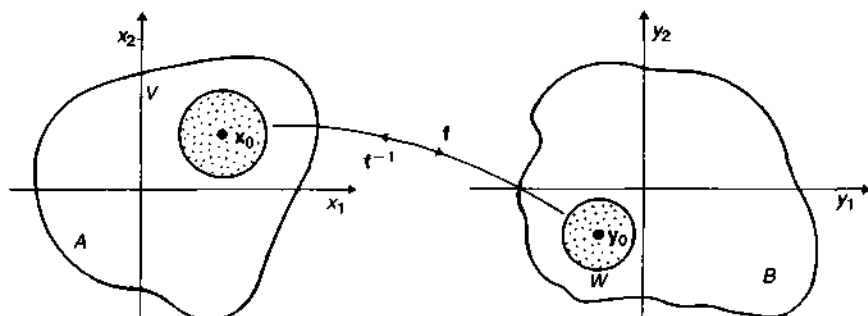



Fig. 7.15

L'idea è di sfruttare le informazioni in nostro possesso riguardanti le trasformazioni lineari, *"linearizzando"* l'equazione intorno ad $\mathbf{x}_0$.

Ciò vuol dire: considerare l'equazione (2.15) nella forma equivalente

$$\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0) = \mathbf{y} - \mathbf{y}_0 \ ;$$

approssimare l'incremento a primo membro con il differenziale di $\mathbf{f}$ in $\mathbf{x}_0$; cioè:

$$\mathbf{J_f}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) \simeq \mathbf{y} - \mathbf{y}_0$$

che è un sistema *lineare*.

Se supponiamo $\mathbf{J_f}(\mathbf{x}_0)$ non singolare abbiamo:

$$\mathbf{x} \simeq \mathbf{x}_0 + (\mathbf{J_f}(\mathbf{x}_0))^{-1}(\mathbf{y} - \mathbf{y}_0) \ . \tag{2.16}$$

Possiamo allora considerare il valore a secondo membro della (2.16) come un valore approssimato di $\mathbf{x}$ e iterare il procedimento definendo la successione

$$\mathbf{x}_{k+1} = \mathbf{x}_k + (\mathbf{J_f}(\mathbf{x}_k))^{-1}(\mathbf{y} - \mathbf{f}(\mathbf{x}_k)) \ ; \tag{2.17}$$

naturalmente occorre assicurarsi che $\mathbf{J_f}(\mathbf{x}_k)$ ad ogni passo sia non singolare.

Se si riesce a dimostrare che $\mathbf{x}_k \to \mathbf{x}$ per $k \to +\infty$, passando al limite nella (2.17) e supponendo $\mathbf{f}$ e $\mathbf{J_f}$ continue si ottiene

$$\mathbf{x} = \mathbf{x} + (\mathbf{J_f}(\mathbf{x}))^{-1}(\mathbf{y} - \mathbf{f}(\mathbf{x}))$$

ovvero $\mathbf{y} = f(\mathbf{x})$.

Quello che abbiamo descritto sommariamente non è altro che la generalizzazione del metodo di Newton per la soluzione delle equazioni del tipo $f(x) = 0$.

Nel caso presente tutto "funziona" bene se $\mathbf{f}$ è di classe C^1 e se $\mathbf{y}$ è abbastanza vicino ad $\mathbf{y}_0$ (da qui il carattere locale del procedimento). Questo è quanto affermato nel prossimo importante teorema.

▪ **Teorema 2.3** - (di inversione locale).
Sia $\mathbf{f} : \mathbb{R}^n \supseteq A \to \mathbb{R}^n$, *A aperto. Se:*

i) $\mathbf{f} \in C^1(A)$

ii) $\mathbf{x}_0$ è *un punto di A tale che* $\mathbf{J_f}(\mathbf{x}_0)$ *è non singolare;*

allora esistono un intorno V di $\mathbf{x}_0$ *ed un intorno W di* $\mathbf{y}_0 = \mathbf{f}(\mathbf{x}_0)$ *tali che:*

a) $\mathbf{f}$ *è una corrispondenza biunivoca tra V e W;*

b) *detta* $\mathbf{g} : W \to V$ *la funzione inversa di* $\mathbf{f}$ *(ristretta a V),* $\mathbf{g} \in C^1(W)$ *e se* $\mathbf{x} = \mathbf{g}(\mathbf{y})$, *vale la formula*

$$\mathbf{J_g}(\mathbf{y}) = (\mathbf{J_f}(\mathbf{x}))^{-1} . \tag{2.18}$$

La formula (2.18) generalizza l'analoga formula per la derivata delle funzioni inverse di funzioni reali di variabile reale.

Premettiamo alla dimostrazione alcune osservazioni ed esempi.

Osservazione 2.1 - Supponiamo che: $\mathbf{f} \in C^1(A), A$ aperto di $\mathbb{R}^n$; $B = \mathbf{f}(A)$ e $\forall \mathbf{x} \in A$, $\mathbf{J_f}(\mathbf{x})$ sia non singolare; ciò significa che , per ogni coppia di punti $\mathbf{x}$ e $\mathbf{y} = \mathbf{f}(\mathbf{x})$ si possono trovare due intorni $V_\mathbf{x}$ e $W_\mathbf{y}$ in corrispondenza biunivoca tramite $\mathbf{f}$. Si dice allora che **f è un diffeomorfismo locale** (di classe C^1) tra A e B.

Naturalmente ciò non significa che $\mathbf{f}$ sia $1 \leftrightarrow 1$ da A in B, come mostra il seguente esempio:

***Esempio* 2.16** - Sia $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ definita da $\mathbf{f}(x,y) = (e^x \cos y, e^x \sin y)$.

Chiamiamo (ξ, η) le variabili nel codominio e consideriamo le equazioni

$$\begin{cases} e^x \cos y = \xi \\ e^x \sin y = \eta . \end{cases}$$

Si ha:

$$\mathbf{J_f}(x,y) = \begin{pmatrix} e^x \cos y & -e^x \sin y \\ e^x \sin y & e^x \cos y \end{pmatrix}$$

e $\det \mathbf{J_f}(x,y) = e^{2x} \neq 0 \qquad \forall (x,y) \in \mathbb{R}^2$.

Dunque $\mathbf{f}$ è un diffeomorfismo locale (di classe C^∞) da $\mathbb{R}^2$ in $\mathbb{R}^2$. Non è però una corrispondenza biunivoca da $\mathbb{R}^2$ in $\mathbb{R}^2$ in quanto

$$\mathbf{f}(0,0) = \mathbf{f}(0, 2k\pi) \quad , \quad \forall k \in \mathbb{Z} .$$

Osservazione 2.2 - Se $|\mathbf{J_f}(\mathbf{x}_0)| := \det \mathbf{J_f}(\mathbf{x}_0) = 0$, non si può affermare nulla in generale riguardo alla possibilità di risolvere l'equazione $\mathbf{f}(\mathbf{x}) = \mathbf{y}$, già nel caso $n = 1$.

Se comunque la funzione inversa $\mathbf{g}$ esiste (localmente) *non* può essere differenziabile; se lo fosse, avremmo, per la (2.18) (posto $\mathbf{y}_0 = \mathbf{f}(\mathbf{x}_0)$).

$$|\mathbf{J}_{\mathbf{g}}(\mathbf{y}_0)| \cdot |\mathbf{J}_{\mathbf{f}}(\mathbf{x}_0)| = 1$$

che è assurdo, essendo il primo membro dell'equazione uguale a zero.

Osservazione 2.3 - La (2.18) assegna la seguente formula per le derivate parziali di $\mathbf{g}$:

$$\frac{\partial g_j}{\partial y_k}(\mathbf{y}) = \frac{1}{|\mathbf{J}_{\mathbf{f}}(\mathbf{x})|} \cdot F_{kj}(\mathbf{x}) \qquad (\mathbf{x} = \mathbf{g}(\mathbf{y})), \tag{2.19}$$

dove F_{kj} è il complemento algebrico di $\dfrac{\partial f_k}{\partial x_j}$ nella matrice $\mathbf{J}_{\mathbf{f}}(\mathbf{x})$.

Nel caso $n = 2$, risolvendo

$$\begin{cases} f_1(x,y) = \xi \\ f_2(x,y) = \eta \end{cases}$$

si ottiene $g_1(\xi,\eta) = x, g_2(\xi,\eta) = y$ per la funzione inversa $\mathbf{g}$; la (2.19) in questo caso dà

$$\frac{\partial g_1}{\partial \xi} = \frac{1}{|\mathbf{J}_{\mathbf{f}}|}\frac{\partial f_2}{\partial y} \qquad \frac{\partial g_1}{\partial \eta} = \frac{-1}{|\mathbf{J}_{\mathbf{f}}|}\frac{\partial f_1}{\partial y}$$

$$\frac{\partial g_2}{\partial \xi} = -\frac{1}{|\mathbf{J}_{\mathbf{f}}|}\frac{\partial f_2}{\partial x} \qquad \frac{\partial g_2}{\partial \eta} = \frac{1}{|\mathbf{J}_{\mathbf{f}}|}\frac{\partial f_1}{\partial x}\ ,$$

dove occorre naturalmente pensare x ed y come funzioni di ξ ed η. Per la funzione dell'esempio 2.16 queste formule danno:

$$\frac{\partial g_1}{\partial \xi}(\xi,\eta) = e^{-2x}e^x \cos y = e^{-x}\cos y\ ;\ \frac{\partial g_1}{\partial \eta}(\xi,\eta) = -e^{-2x}(-e^x \sin y) = e^{-x}\sin y$$

$$\frac{\partial g_2}{\partial \xi}(\xi,\eta) = -e^{-2x}e^x \sin y = -e^{-x}\sin y\ ;\ \frac{\partial g_2}{\partial \eta}(\xi,\eta) = e^{-2x}(e^x \cos y) = e^{-x}\cos y.$$

In particolare notando che $(x,y) = (0,0)$ corrisponde (localmente) a $(\xi,\eta) = (1,0)$ abbiamo

$$\mathbf{J}_{\mathbf{g}}(1,0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Prima di iniziare la dimostrazione del teorema 2.3 osserviamo che si può provare una forma più debole del teorema del valor medio per funzioni vettoriali.

Sia infatti $\mathbf{f} : \mathbb{R}^n \supseteq A \to \mathbb{R}^m$, di classe C^1; se il segmento $[\mathbf{x}, \mathbf{x}+\mathbf{h}]$ è contenuto in A, applicando il teorema del valor medio ad ogni componente di $\mathbf{f}$ (che è scalare), si deduce che, per ogni $j = 1, \dots, n$, esiste $\theta_j, 0 < \theta_j < 1$, tale che

$$f_j(\mathbf{x}+\mathbf{h}) - f_j(\mathbf{x}) = < \nabla f_j(\mathbf{x}+\theta_j \mathbf{h}), \mathbf{h} > \quad . \tag{2.20}$$

Se poniamo $\alpha_j = \max\limits_{[\mathbf{x},\mathbf{x}+\mathbf{h}]} \|\nabla f_j\|$, dalla (2.20) e dalla disuguaglianza di Cauchy–Schwarz, otteniamo

$$|f_j(\mathbf{x}+\mathbf{h}) - f_j(\mathbf{x})| \leq \|\nabla f_j(\mathbf{x}+\theta_j \mathbf{h})\| \cdot \|\mathbf{h}\| \leq \alpha_j \|\mathbf{h}\|;$$

infine, se $\alpha = \max\{\alpha_1, \alpha_2, \dots, \alpha_n\}$:

$$\begin{aligned} \|\mathbf{f}(\mathbf{x}+\mathbf{h}) - \mathbf{f}(\mathbf{x})\| &= \left(\sum_{j=1}^{n} (f_j(\mathbf{x}+\mathbf{h}) - f_j(\mathbf{x}))^2\right)^{1/2} \leq \\ &\leq \left(\sum_{j=1}^{n} \alpha_j^2 \|\mathbf{h}\|^2\right)^{1/2} \leq \sqrt{n}\alpha \|\mathbf{h}\| \, . \end{aligned} \tag{2.21}$$

La (2.21) è la generalizzazione (in forma più debole per via della disuguaglianza che vi compare) del teorema del valor medio per funzioni vettoriali.

Dimostrazione del teorema di inversione locale - Poiché si tratta di una dimostrazione piuttosto complessa la eseguiremo in vari passi.

1) Riduzione al caso $\mathbf{x}_0 = \mathbf{0}$, $\mathbf{y}_0 = \mathbf{0}, \mathbf{J}_\mathbf{f}(\mathbf{x}_0) = \mathbf{I}_n$ (matrice identità in $\mathbb{R}^n$), ovvero, riduzione dell'equazione $\mathbf{f}(\mathbf{x}) = \mathbf{y}$ ad un'equazione $\mathbf{F}(\mathbf{u}) = \mathbf{w}$, con $\mathbf{u}$ e $\mathbf{w}$ in un intorno di $\mathbf{0}, \mathbf{F}(\mathbf{0}) = \mathbf{0}$ e $\mathbf{J}_\mathbf{F}(\mathbf{0}) = \mathbf{I}_n$.

L'equazione $\mathbf{f}(\mathbf{x}) = \mathbf{y}$ è equivalente a

$$\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0) = \mathbf{y} - \mathbf{y}_0. \tag{2.22}$$

Poniamo $\mathbf{x} = \mathbf{u} + \mathbf{x}_0$ e moltiplichiamo entrambi i membri della (2.22) per l'inversa di $\mathbf{J}_\mathbf{f}(\mathbf{x}_0)$; si ottiene l'equazione equivalente

$$(\mathbf{J}_\mathbf{f}(\mathbf{x}_0))^{-1}[\mathbf{f}(\mathbf{u}+\mathbf{x}_0) - \mathbf{f}(\mathbf{x}_0)] = (\mathbf{J}_\mathbf{f}(\mathbf{x}_0))^{-1}(\mathbf{y} - \mathbf{y}_0). \tag{2.23}$$

Poniamo ora

$$\mathbf{F}(\mathbf{u}) = (\mathbf{J}_\mathbf{f}(\mathbf{x}_0))^{-1}[\mathbf{f}(\mathbf{u}+\mathbf{x}_0) - \mathbf{f}(\mathbf{x}_0)]$$

e

$$\mathbf{w} = (\mathbf{J}_\mathbf{f}(\mathbf{x}_0))^{-1}(\mathbf{y} - \mathbf{y}_0) \ ;$$

la (2.23) si scrive

$$\mathbf{F}(\mathbf{u}) = \mathbf{w}.$$

Si ha: $\mathbf{F}(\mathbf{0}) = \mathbf{0}$, $\mathbf{F}$ soddisfa *i)* nell'aperto $A_0 = \{\mathbf{u} : \mathbf{u} + \mathbf{x}_0 \in A\}$ e *ii)* con $\mathbf{x}_0 = \mathbf{0}$ ed inoltre

$$\mathbf{J}_\mathbf{F}(\mathbf{0}) = (\mathbf{J}_\mathbf{f}(\mathbf{x}_0))^{-1} \cdot \mathbf{J}_\mathbf{f}(\mathbf{x}_0) = \mathbf{I}_n \ . \tag{2.24}$$

2) Facciamo vedere che $\mathbf{F}$ è iniettiva in un intorno di $\mathbf{0}$.
Poniamo

$$\boldsymbol{\phi}(\mathbf{u}) = \mathbf{u} - F(\mathbf{u}) \quad , \quad \mathbf{u} \in A_0,$$

e sia η, $0 < \eta < 1$. Allora $\boldsymbol{\phi} \in C^1(A_0)$ e $\boldsymbol{\phi}(0) = \mathbf{0}, \mathbf{J}_{\boldsymbol{\phi}}(\mathbf{0}) = \mathbf{O}$ (= matrice nulla).

Essendo tutte le derivate di $\boldsymbol{\phi}$ nulle in $\mathbf{0}$ ed ivi continue, esiste una palla B_r centrata in $\mathbf{0}$ e di raggio r, tale che, per ogni $j = 1, \ldots, n$, $\|\nabla \phi_j\| < \frac{1}{\sqrt{n}}\eta$.

Dunque, in base alla (2.21) possiamo scrivere, per ogni coppia di punti $\mathbf{u}_1, \mathbf{u}_2 \in B_r$:

$$\|\boldsymbol{\phi}(\mathbf{u}_1) - \boldsymbol{\phi}(\mathbf{u}_2)\| \leq \eta\|\mathbf{u}_1 - \mathbf{u}_2\| ; \tag{2.25}$$

essendo

$$\|\boldsymbol{\phi}(\mathbf{u}_1) - \boldsymbol{\phi}(\mathbf{u}_2)\| = \|\mathbf{u}_1 - \mathbf{F}(\mathbf{u}_1) - \mathbf{u}_2 + \mathbf{F}(\mathbf{u}_2)\| \geq \|\mathbf{u}_1 - \mathbf{u}_2\| - \|\mathbf{F}(\mathbf{u}_1) - \mathbf{F}(\mathbf{u}_2)\|$$

dalla (2.25) si ricava

$$\|\mathbf{F}(\mathbf{u}_1) - \mathbf{F}(\mathbf{u}_2)\| \geq (1 - \eta)\|\mathbf{u}_1 - \mathbf{u}_2\| \tag{2.26}$$

da cui si deduce che $\mathbf{F}(\mathbf{u}_1) = \mathbf{F}(\mathbf{u}_2)$ implica $\mathbf{u}_1 = \mathbf{u}_2$ e cioè che $\mathbf{F}$ è iniettiva. Si può inoltre scegliere r (eventualmente diminuendolo) in modo che $\mathbf{J}_{\mathbf{F}}(\mathbf{u})$ sia non singolare per ogni $\mathbf{u} \in B_r$.

Si noti inoltre che se $\mathbf{G}$ è l'inversa di $\mathbf{F}_{|B_r}$ la (2.26), posto $\mathbf{w}_1 = \mathbf{F}(\mathbf{u}_1)$ e $\mathbf{w}_2 = \mathbf{F}(\mathbf{u}_2)$, si può riscrivere nella forma

$$\|\mathbf{G}(\mathbf{w}_1) - \mathbf{G}(\mathbf{w}_2)\| \leq (1 - \eta)^{-1}\|\mathbf{w}_1 - \mathbf{w}_2\| \tag{2.26'}$$

che mostra la continuità di $\mathbf{G}$.

3) Occorre dimostrare ora che $\mathbf{F}(B_r)$ contiene un intorno W dell'origine.
La tesi a) del teorema sarà allora verificata con $V = \mathbf{G}(W)$.

Proviamo che come intorno dell'origine si può scegliere B_α con $\alpha = (1 - \eta)\frac{r}{2}$, ovvero la palla di raggio $(1 - \eta)\frac{r}{2}$, dove η è il numero fissato nel passo 2).

Infatti, sia $\mathbf{w} \in B_\alpha$; dobbiamo trovare $\mathbf{u} \in B_r$ tale che $\mathbf{F}(\mathbf{u}) = \mathbf{w}$. A tale scopo definiamo per ricorrenza la successione seguente:

$$\mathbf{u}_0 = \mathbf{0} \ , \ \mathbf{u}_{n+1} = \mathbf{w} + \boldsymbol{\phi}(\mathbf{u}_n) \tag{2.27}$$

e mostriamo che essa è contenuta in B_r, che è di Cauchy e che converge ad un punto $\mathbf{u}$ di B_r. Passando poi al limite nella (2.27), essendo $\boldsymbol{\phi}$ continua, si ha:

$$\mathbf{u} = \mathbf{w} + \boldsymbol{\phi}(\mathbf{u}) \quad \text{che equivale a} \quad \mathbf{F}(\mathbf{u}) = \mathbf{w} \ .$$

Procediamo per induzione; evidentemente $\mathbf{u}_0 \in B_r$ e siano $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n \in B_r$. Allora, per la disuguaglianza triangolare

$$\|\mathbf{u}_{n+1}\| \leq \|\mathbf{u}_{n+1} - \mathbf{u}_n\| + \|\mathbf{u}_n - \mathbf{u}_{n-1}\| + \ldots + \|\mathbf{u}_1 - \mathbf{u}_0\|; \tag{2.28}$$

ma, per la (2.25), per ogni $k = 1, \ldots, n$, si ha:

$$\|\mathbf{u}_{k+1} - \mathbf{u}_k\| = \|\boldsymbol{\phi}(\mathbf{u}_k) - \boldsymbol{\phi}(\mathbf{u}_{k-1})\| \leq \eta\|\mathbf{u}_k - \mathbf{u}_{k-1}\|$$

che, iterata da k fino a 1, dà, essendo $||\mathbf{w}|| < (1-\eta)\frac{r}{2}$

$$||\mathbf{u}_{k+1} - \mathbf{u}_k|| \leq \eta^k ||\mathbf{u}_1 - \mathbf{u}_0|| = \eta^k ||\mathbf{w}|| < \eta^k (1-\eta)\frac{r}{2}. \tag{2.29}$$

Usando la (2.29) dalla (2.28) si ricava

$$||\mathbf{u}_{n+1}|| \leq (1-\eta)\frac{r}{2}\sum_{k=1}^{n} \eta^k = \frac{r}{2}(1-\eta^{n+1}) < \frac{r}{2}. \tag{2.30}$$

Perciò anche $\mathbf{u}_{n+1} \in B_r$ e la successione (2.27) è ben definita.

Per vedere che è di Cauchy, dalla (2.29) e dalla disuguaglianza triangolare si ottiene, per $m > n$:

$$||\mathbf{u}_m - \mathbf{u}_n|| \leq ||\mathbf{u}_m - \mathbf{u}_{m-1}|| + \cdots + ||\mathbf{u}_{n+1} - \mathbf{u}_n|| \leq (1-\eta)\frac{r}{2}\sum_{k=n}^{m} \eta^k =$$

$$= \eta^n (1-\eta)\frac{r}{2}\sum_{k=0}^{m-n} \eta^k = \eta^n (1-\eta^{m-n+1})\frac{r}{2} < \eta^n \frac{r}{2}.$$

Essendo $\eta < 1$, $||\mathbf{u}_m - \mathbf{u}_n|| \to 0$ per $n \to +\infty$; di conseguenza $\{\mathbf{u}_n\}$ è di Cauchy ed esiste un punto $\mathbf{u}$ tale che $\mathbf{u}_n \to \mathbf{u}$.

Dalla (2.30), poiché $||\mathbf{u}_n|| \to ||\mathbf{u}||$, si ottiene che $\mathbf{u} \in \overline{B}_{r/2} \subset B_r$.

4) Rimane da dimostrare che $G \in C^1(W)$ e che vale la formula (2.18).

Siano $\mathbf{w}$ e $\mathbf{w} + \mathbf{k}$ punti di $W (\mathbf{k} \neq \mathbf{0})$; poniamo $\mathbf{u} = \mathbf{G}(\mathbf{w})$ e $\mathbf{h} = \mathbf{G}(\mathbf{w}+\mathbf{k}) - \mathbf{G}(\mathbf{w})$. Allora $\mathbf{u} + \mathbf{h} = \mathbf{G}(\mathbf{w}+\mathbf{k})$ che equivale a $\mathbf{F}(\mathbf{u}+\mathbf{h}) = \mathbf{w}+\mathbf{k}$ ovvero a $\mathbf{k} = \mathbf{F}(\mathbf{u}+\mathbf{h}) - \mathbf{F}(\mathbf{u})$. Poiché $\mathbf{F}$ è di classe C^1, si può scrivere

$$\mathbf{k} = \mathbf{J_F}(\mathbf{u})\mathbf{h} + \mathbf{E}(\mathbf{h}) \text{ dove } \frac{||\mathbf{E}(\mathbf{h})||}{||\mathbf{h}||} \to 0 \text{ per } \mathbf{h} \to \mathbf{0};$$

essendo $\mathbf{J_F}(\mathbf{u})$ non singolare per $\mathbf{u} \in B_r$ (vedi punto 1) abbiamo allora:

$$\mathbf{G}(\mathbf{w}+\mathbf{k}) - \mathbf{G}(\mathbf{w}) = \mathbf{h} = (\mathbf{J_F}(\mathbf{u}))^{-1}\mathbf{k} + (\mathbf{J_F}(\mathbf{u}))^{-1}\mathbf{E}(\mathbf{h}). \tag{2.31}$$

Ricordiamo ora che, dalla (2.26') si deduce $||\mathbf{h}|| \leq (1-\eta)^{-1}||\mathbf{k}||$ (dunque $\mathbf{h} \to \mathbf{0}$ se $\mathbf{k} \to \mathbf{0}$); inoltre è facile verificare che

$$||(\mathbf{J_F}(\mathbf{u}))^{-1}\mathbf{E}(\mathbf{h})|| \leq nc||\mathbf{E}(\mathbf{h})|| \tag{2.32}$$

dove c è una costante che maggiora in modulo tutti gli elementi della matrice $(\mathbf{J_F}(\mathbf{u}))^{-1}$. Si può allora scrivere

$$\frac{||(\mathbf{J_F}(\mathbf{u}))^{-1}\mathbf{E}(\mathbf{h})||}{||\mathbf{k}||} \leq \frac{nc}{1-\eta}\frac{||\mathbf{E}(\mathbf{h})||}{||\mathbf{h}||} \to 0 \text{ per } \mathbf{k} \to \mathbf{0}.$$

Si conclude dalla (2.31) che $\mathbf{G}$ è differenziabile in $\mathbf{w}$ con matrice Jacobiana $(\mathbf{J_F}(\mathbf{u}))^{-1}$ dove $\mathbf{u} = \mathbf{G}(\mathbf{w})$. La (2.18) è quindi dimostrata per $\mathbf{G}$ e con essa il fatto che $\mathbf{G}$ sia di classe $C^1(W)$; è poi facile reinterpretare le conclusioni per $\mathbf{f}$ e la sua inversa $\mathbf{g}$. □

Esercizi

1. Scrivere la matrice Jacobiara per le seguenti funzioni vettoriali in un punto generico dello spazio.

a) $\mathbf{f}(x,y) = (ye^{x^2}, \sin(x+y))$

b) $\mathbf{f}(x,y,z) = \left(\dfrac{z}{x^2+1}, y-3z, 4x^2+e^y, z^3-x\right)$

2. Scrivere esplicitamente $\dfrac{\partial \tilde{f}}{\partial u}$, $\dfrac{\partial \tilde{f}}{\partial v}$ se $\tilde{f}(u,v) = f(x(u,v), y(u,v))$.

3. Interpretare geometricamente l'immagine delle seguenti funzioni da $\mathbb{R} \supseteq I$ in $\mathbb{R}^2$ o $\mathbb{R}^3$ oppure da $\mathbb{R}^2 \supseteq T$ in $\mathbb{R}^3$.

a) $\mathbf{r}(t) = (x_0 + a\cos t, y_0 + b\sin t) \quad a,b \in \mathbb{R}_+ \ , \ t \in [0, 2\pi)$.

b) $\mathbf{r}(t) = (t, t^2) \quad t \in \mathbb{R}$

c) $\mathbf{r}(t) = (x_0 + a\cos t, y_0 + b\sin t, kt) \quad a,b \in \mathbb{R}_+ \ , \ t \in \mathbb{R}$.

d) $\mathbf{r}(u,v) = (R\cos u, R\sin u, v) \quad u,v \in [0,2\pi) \times \mathbb{R}$

e) $\mathbf{r}(u,v) = (R\sin u\cos v, R\sin u\sin v, R\cos u) \quad R \in \mathbb{R}_+, \ (u,v) \in [0,\pi] \times [0,2\pi)$

f) $\mathbf{r}(u,v) = (a\sin u\cos v, b\sin u\sin v, c\cos u)\ , \ a,b,c \in \mathbb{R}_+, (u,v) \in [0,\pi] \times [0,2\pi)$

g) $\mathbf{r}(u,v) = (u, v, u^2+v^2)$

h) $\mathbf{r}(u,v) = (u, v, \sqrt{u^2+v^2})$.

Per ciascuna di esse, calcolare $\mathbf{r}'(t)$ o $\mathbf{r}_u(u,v)$ e $\mathbf{r}_v(u,v)$.

4. Siano $f\colon \mathbb{R}^2 \supseteq A \to \mathbb{R}$, $(x_0,y_0) \in A$ e $z_0 = f(x_0,y_0)$.
Se f è derivabile lungo ogni direzione in (x_0,y_0) diciamo che f "ha la proprietà $\mathcal{P}$ relativamente a (x_0,y_0)" se tutte le tangenti alla sezione del grafico di f con un piano ortogonale al piano $z=0$ e passante per $(x_0,y_0,0)$ sono complanari.
Vero o falso:

a) Se f è differenziabile in (x_0,y_0), allora ha la proprietà $\mathcal{P}$ relativamente a (x_0,y_0).

b) Se f ha la proprietà $\mathcal{P}$ relativamente ad (x_0,y_0) è differenziabile in (x_0,y_0)

c)** Se f ha la proprietà $\mathcal{P}$ relativamente a (x_0,y_0) allora vale la formula (1.3):
$D_{\mathbf{v}}f(x_0,y_0) = <\nabla f(x_0,y_0), \mathbf{v}>$ per ogni versore $\mathbf{v} \in \mathbb{R}^2$.

[Suggerimento: b) si veda l'esercizio 11 della sezione 1.
c) se un piano ortogonale al piano $z=0$ ha per intersezione con quest'ultimo la retta $x = x_0 + tv_1$, $y = y_0 + tv_2$ dove $\mathbf{v} = (v_1,v_2)$ è un versore, allora il vettore $(v_1, v_2, D_{\mathbf{v}}f(x_0,y_0))$ è tangente alla sezione nel punto (x_0,y_0,z_0); se tutti questi vettori giacciono su uno stesso piano per ogni $\mathbf{v}$, in questo piano due vettori linearmente indipendenti sono dati da $(1,0,f_x(x_0,y_0))$ e $(0,1,f_y(x_0,y_0))\ldots$].

5. Sia $f: \mathbb{R}^3 \setminus \{0,0,0\} \to \mathbb{R}$. Poniamo $\tilde{f}(\rho,\theta,t) = f(\rho\cos\theta, \rho\sin\theta, t)$ (cambiamento in coordinate cilindriche) e $\overline{f}(\rho,\theta,\psi) = f(\rho\sin\psi\cos\theta, \rho\sin\psi\sin\theta, \rho\cos\psi)$, (cambiamento in coordinate polari).
Esprimere le derivate parziali di $\tilde{f}$ e di $\overline{f}$ in termini delle derivate parziali di f.

6. Verificare che data $f \in C^2(A)$, $A \subseteq \mathbb{R}^n$ aperto, si ha:

a) $\operatorname{div} \operatorname{grad} f = \Delta f$

b) $\operatorname{tr} \mathbf{H}_f = \Delta f$. (tr sta per "traccia") (*).

Utilizzare la b) per dimostrare che se $\mathbf{M}$ è una matrice ortogonale; posto $\tilde{f}(y) = f(\mathbf{M}x)$ si ha $\Delta \tilde{f}(y) = \sum_{j=i}^{n} \frac{\partial^2 \tilde{f}}{\partial y_j^2}(y)$ e cioè l'operatore di Laplace ha la stessa forma in qualunque sistema di coordinate ortogonali, ovvero è un invariante per trasformazioni ortogonali nello spazio. (Ricordare la (2.9) e che se $\mathbf{M}$ è non singolare $\operatorname{tr}(\mathbf{M}^{-1}\mathbf{N}\mathbf{M}) = \operatorname{tr}\mathbf{N}$).

7.* a) Dimostrare che, per una funzione f differenziabile in $\mathbb{R}^n$, ∇f è un invariante per trasformazioni ortogonali nello spazio (ovvero se $\mathbf{M}$ è una matrice ortogonale e $\tilde{f}(y) = f(\mathbf{M}x)$ allora $\nabla \tilde{f}(y) = (\tilde{f}_{y_1}(y), \tilde{f}_{y_2}(y), \ldots, \tilde{f}_{y_n}(y))$.

8.* Se $\mathbf{f} : \mathbb{R}^3 \to \mathbb{R}^3$ è differenziabile, dimostrare che divergenza e rotore di $\mathbf{f}$ sono invarianti per trasformazioni ortogonali dello spazio.

9. Sia $\mathbf{r} = (x, y, z)$ e $r = \|\mathbf{r}\|$; posto $\mathbf{f} = \frac{\mathbf{r}}{r^3}$ calcolare $\operatorname{rot}\mathbf{f}$ e $\operatorname{div}\mathbf{f}$.

10. Sia $f(t) = F(x(t), y(t))$ dove

$$F(x, y) = x^2(y^3 + y - 4) + y(3x^2 + 2x) .$$

Calcolare $f'(0)$ e $f''(0)$.

11. Sia $f : \mathbb{R}^2 \to \mathbb{R}$ di classe $C^1(\mathbb{R}^2)$. Definiamo la seguente successione per ricorrenza:

$$F_0(x, y) = f(x, y)$$

$$F_n(x, y) = f(F_{n-1}(x, y), F_{n-1}(x, y)) \quad n \geq 1 .$$

Dimostrare che

$$\frac{\partial F_n}{\partial x} \cdot \frac{\partial F_m}{\partial y} - \frac{\partial F_n}{\partial y} \cdot \frac{\partial F_m}{\partial x} = 0 \quad \forall (x, y) \in \mathbb{R}^2, \ \forall m, n \in \mathbb{N} .$$

12.* Sia $f \in C^2(\mathbb{R})$ scalare. Dimostrare che se $f = f(x, t)$ soddisfa l'equazione (della corda vibrante)

$$\frac{\partial^2 f}{\partial t^2} - c^2 \frac{\partial^2 f}{\partial x^2} = 0 \qquad (c \in \mathbb{R}_+)$$

(*) Se $\mathbf{M}$ è una matrice quadrata di ordine n, $\operatorname{tr}\mathbf{M} := \sum_{i=1}^{n} m_{ii}$.

allora, posto $\xi = ct - x$, $\eta = ct + x$, la funzione $\tilde{f}(\xi, \eta) = f\left(\frac{\eta - \xi}{2}, \frac{\eta + \xi}{2c}\right)$ soddisfa l'equazione

$$\frac{\partial^2 \tilde{f}}{\partial \xi \partial \eta} = 0 .$$

Utilizzare l'esercizio 1.15 per dedurre che una generica soluzione della corda vibrante è della forma

$$f(x, y) = \varphi(ct - x) + \psi(ct + x) .$$

Interpretare la soluzione trovata come sovrapposizione di due onde progressive, una che si muove con velocità c nella direzione positiva dell'asse x, l'altra con la stessa velocità in direzione opposta.

[Esaminare il comportamento dei grafici delle funzioni φ e ψ considerando t come parametro].

13.* a) Verificare che per una funzione $f \in C^2(\mathbf{R}^2)$ in coordinate polari si ha, per $\rho \neq 0$:

$$\Delta f = \tilde{f}_{\rho\rho} + \frac{1}{\rho}\tilde{f}_\rho + \frac{1}{\rho^2}\tilde{f}_{\theta\theta} \quad (\tilde{f}(\rho, \theta) = f(\rho\cos\theta, \rho\sin\theta)) .$$

b) Utilizzando la formula in a) dedurre che, se $f \in C^2(\mathbf{R}^3)$ in coordinate cilindriche si ha, per $\rho \neq 0$:

$$\Delta f = \tilde{f}_{\rho\rho} + \frac{1}{\rho}\tilde{f}_\rho + \frac{1}{\rho^2}\tilde{f}_{\theta\theta} + \tilde{f}_{zz} \quad (\tilde{f}(\rho, \theta, z) = f(\rho\cos\theta, \rho\sin\theta, z)) .$$

c) Ponendo in b) $z = r\cos\psi$, $\rho = r\sin\psi$ ($\psi \in [0, \pi]$) si passa in coordinate polari nello spazio; posto $\overline{f}(r, \theta, \psi) = \tilde{f}(r\sin\psi, \theta, r\cos\psi)$ dedurre, per $\rho \neq 0$, $\psi \neq 0, \pi$ la seguente formula per l'operatore di Laplace in coordinate polari nello spazio:

$$\Delta f = \overline{f}_{rr} + \frac{2}{r}\overline{f}_r + \frac{1}{r^2}\overline{f}_{\psi\psi} + \frac{\cos\psi}{r^2(\sin\psi)^2}\overline{f}_\psi .$$

14. Dimostrare che la proprietà di invarianza formale del differenziale primo continua a valere.

15. Stabilire quali tra le seguenti funzioni complesse sono olomorfe nel loro dominio.

a) $\mathrm{Re}(z)$

b) $\mathrm{Im}(z)$

c) $|z|$

d) $a_0 + a_1 z + a_2 z^2 + \ldots + a_n z^n \quad (a_j \in \mathbb{C})$

e) $\dfrac{az + b}{cz + d} \quad (a, b, c, d \in \mathbb{C})$.

16. Sia f olomorfa in $A \subseteq \mathbb{C}$, A aperto, $f(z) = u(x,y) + i\, v(x,y)$.
Verificare che

a) $|f'(z)|^2 = \dfrac{\partial(u,v)}{\partial(x,y)}$.

b) $< \nabla u, \nabla v >= 0$.

17. Scrivere le condizioni di Cauchy–Riemann in coordinate polari.

18. Costruire esempi di funzioni armoniche in aperti del piano.
Verificare che le funzioni $r^n \cos n\theta$, $r^n \sin n\theta$ $(n \in \mathbb{N})$ sono armoniche in $\mathbb{R}^2$; esse sono dette *armoniche elementari*.

19. Sia $\alpha < 0$ e $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ definita da

$$f_1(x,y) = \begin{cases} x(x^2+y^2)^\alpha & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases} \quad f_2(x,y) = \begin{cases} y(x^2+y^2)^\alpha & (x,y) \neq (0,0) \\ 0 & (x,y) = (0,0) \end{cases}.$$

Per quali α $\mathbf{f}$ è invertibile in un intorno dell'origine?

20. Verificare l'invertibilità locale per le trasformazioni in coordinate polari nel piano e nello spazio.

21. Dimostrare che $f : \mathbb{C} \to \mathbb{C}$ è localmente invertibile se $f'(z) \neq 0$. Dare la formula per la derivata della funzione inversa.

22. Sia $f : \mathbb{R} \to \mathbb{R}$ di classe C^1; consideriamo la seguente trasformazione

$$\begin{cases} \xi = f(x) \\ \eta = -y + x f(x). \end{cases}$$

Trovare sotto quali condizioni per f, tale trasformazione risulta un diffeomorfismo locale. Calcolare la matrice Jacobiana dell'inversa nel punto $(f(0), -1)$.

3. FUNZIONI IMPLICITE

3.1 Esempi preliminari

Se consideriamo una funzione $g : \mathbb{R}^2 \supseteq A \to \mathbb{R}$, l'equazione

$$g(x,y) = 0 \tag{3.1}$$

individua una relazione in $\mathbb{R}^2$; l'insieme E_0 dei punti del piano che soddisfano la (3.1) costituisce il grafico della relazione stessa. Può succedere che la (3.1) sia "esplicitabile" rispetto ad x o rispetto ad y o, in altri termini, che sia possibile "risolvere" la (3.1) ricavando y in funzione di x oppure x in funzione di y. In tal caso E_0 coincide, parzialmente o interamente, con il grafico di una funzione della forma $y = f(x)$ oppure $x = h(y)$. Tali funzioni si dicono *definite implicitamente* dalla (3.1).

Ad esempio, l'equazione

$$g(x,y) = ax + by + c = 0 \qquad (a \neq 0, b \neq 0)$$

permette di ricavare $y = -\frac{a}{b}x - \frac{c}{b}$ oppure $x = -\frac{b}{a}y - \frac{c}{a}$.

È ben noto che E_0 è una retta.

Dall'equazione

$$xe^y - 3y^2 + 1 = 0$$

è possibile ricavare x in funzione di $y : x = e^{-y}(3y^2 - 1)$ ma non viceversa.

Può anche succedere che un'equazione definisca più di una funzione implicita come nel caso

$$x^2 + y^2 - 1 = 0\ . \tag{3.2}$$

La (3.2), risolta rispetto ad y dà

$$y = +\sqrt{1-x^2} \qquad \text{(il cui grafico è il semicerchio superiore)}$$

e

$$y = -\sqrt{1-x^2} \qquad \text{(il cui grafico è il semicerchio inferiore)}\ .$$

In realtà la (3.2) definisce infinite funzioni implicite della forma $y = \varepsilon(x)\sqrt{1-x^2}$ dove $\varepsilon(x)$ è una qualunque funzione definita in $[-1,1]$ che assume solo i valori $+1$ o -1. La figura 7.16d corrisponde ad $\varepsilon(x) = +1$ per $x \in [-1, -\frac{2}{3}] \cup [0, \frac{2}{3}]$ e $\varepsilon(x) = -1$ per $x \in (-\frac{2}{3}, 0) \cup (\frac{2}{3}, 1]$.

Se però limitiamo l'interesse a funzioni implicite *continue* allora la scelta si restringe alle due sole funzioni $y = \pm\sqrt{1-x^2}$. Se poi imponiamo il passaggio per un punto, diciamo $(0,-1)$, la scelta diventa unica: $y = -\sqrt{1-x^2}$. Osserviamo subito che, in un intorno di un punto x_0 diverso da $0, +1$ e -1 l'equazione (3.2) è risolvibile sia rispetto ad x sia rispetto ad y; non così nell'intorno dei punti $(0,1)$ e $(0,-1)$ dove è possibile ricavare y in funzione di x ma non viceversa, e nell'intorno dei punti $(1,0)$ e $(-1,0)$ dove x è esplicitabile come funzione di y ma non viceversa.

Il fatto che in questi punti la tangente al cerchio unitario sia orizzontale o verticale (rispettivamente) non è casuale, come vedremo in seguito.

Consideriamo ora l'equazione

$$g(x,y) = xe^y + ye^x = 0\ . \tag{3.3}$$

Nessuna manipolazione permette di esplicitare y in funzione di x o viceversa. D'altra parte $g(0,0) = 0$, dunque $(0,0)$ è un punto di E_0 ed è ragionevole porsi il problema se la (3.3), in un intorno di $x = 0$ (o $y = 0$), definisca $y = y(x)$ (o $x = x(y)$).

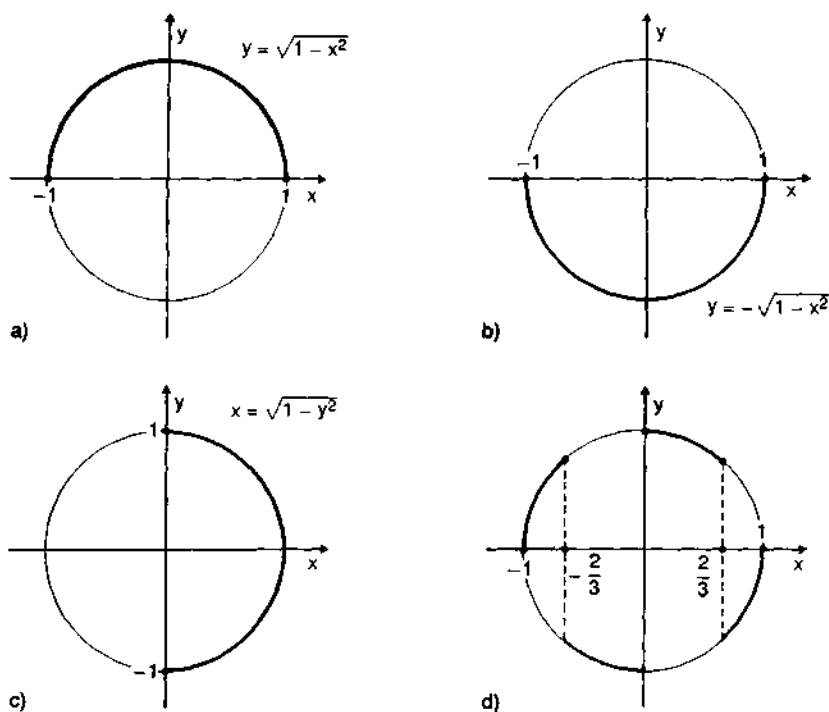



Fig. 7.16 Alcune funzioni implicite definite dall'equazione $x^2 + y^2 - 1 = 0$.

È naturalmente erroneo pensare che *ogni* equazione del tipo $g(x, y) = 0$ definisca sempre implicitamente qualche funzione: $x^2 + y^2 = 0$ definisce ... l'origine $(0, 0)$, mentre $x^2 + y^2 + 1 = 0$ non è soddisfatta da alcuna coppia di numeri reali.

La questione che vogliamo affrontare è:

data l'equazione (3.1), *trarre conclusioni su esistenza, unicità, regolarità delle funzioni implicite da essa definite, dallo studio della funzione g stessa.*

3.2 Il teorema di Dini (*)

Premettiamo una definizione di funzione implicita.

Definizione 3.1 - *Sia* $g : \mathbb{R}^2 \supseteq A \to \mathbb{R}$*; la funzione* $y = f(x), x \in I \subseteq \mathbb{R}$ *(risp.* $x = h(y), y \in I \subseteq \mathbb{R}$*) si dice definita implicitamente dall'equazione* $g(x, y) = 0$ *se*

(*) Ulisse Dini, matematico italiano (1845–1918).

$$\text{graf}(f) \subset A \quad e$$

$$g(x, f(x)) = 0 \quad \forall x \in I \quad (\text{risp. } g(h(y), y) = 0 \quad \forall y \in I) .$$

Il concetto di funzione implicita permette di precisare quello di *funzione algebrica.* Diciamo che f è *algebrica* se f può essere definita implicitamente da un'equazione algebrica, cioè del tipo $g(x, y) = 0$ dove g è un polinomio nelle due variabili x ed y. Le funzioni come esponenziali, logaritmi, funzioni iperboliche e circolari, che non soddisfano alcuna equazione algebrica, si dicono *trascendenti.*

Veniamo ora ad un importante teorema che dà condizioni sufficienti a garantire esistenza ed unicità di una funzione implicita; ci riferiremo, per brevità, solo a funzioni del tipo $y = f(x)$.

■ **Teorema 3.1** - (di Dini). *Sia* $g : \mathbb{R}^2 \supseteq A \to \mathbb{R}$, *A aperto. Supponiamo che:*

i) g *e* g_y *siano continue in* A;

ii) nel punto $(x_0, y_0) \in A$ *si abbia* $g(x_0, y_0) = 0$ *e* $g_y(x_0, y_0) \neq 0$.

Allora esistono un intorno U *di* x_0 *ed un'unica funzione* $f : U \to \mathbb{R}$, *continua in* U, *tale che* $y_0 = f(x_0)$ *e che* $g(x, f(x)) = 0$ *per ogni* $x \in U$. *Se inoltre*

iii) g_x *è continua in* A (dunque $g \in C^1(A)$)

allora $f \in C^1(U)$ *e vale la formula*

$$f'(x) = -\frac{g_x(x, f(x))}{g_y(x, f(x))} \qquad \forall x \in U . \tag{3.4}$$

Dimostrazione - Possiamo sempre supporre $g_y(x_0, y_0) > 0$. Essendo g_y continua in A, per il teorema della permanenza del segno esiste un rettangolo $W = [x_0 - a, x_0 + a] \times$ $\times [y_0 - b, y_0 + b]$ in cui $g_y > 0$. Allora la funzione (della sola y)

$$y \mapsto g(x_0, y) \ , \ y \in [y_0 - b, y_0 + b]$$

è strettamente crescente e continua; inoltre, poiché $g(x_0, y_0) = 0$, si ha:

$$g(x_0, y_0 - b) < 0 \quad \text{e} \quad g(x_0, y_0 + b) > 0. \tag{3.5}$$

Consideriamo ora le funzioni (della sola x):

$$x \mapsto g(x, y_0 - b) \quad \text{e} \quad x \mapsto g(x, y_0 + b),$$

continue in $[x_0 - a, x_0 + a]$. Per le (3.5) e il teorema della permanenza del segno, esiste un intorno di $x_0, (x_0 - \delta, x_0 + \delta)$ in cui, simultaneamente

$$g(x, y_0 - b) < 0 \quad \text{e} \quad g(x, y_0 + b) > 0.$$

Se ora fissiamo $\overline{x} \in (x_0 - \delta, x_0 + \delta)$, la funzione (della sola y)

$$y \mapsto g(\overline{x}, y)$$

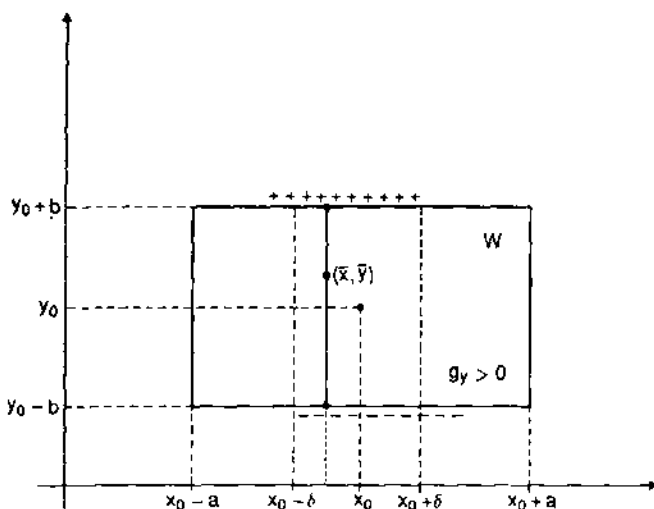



Fig. 7.17

è continua in $[y_0-b, y_0+b]$ e passa da un valore negativo, per $y=y_0-b$, ad uno positivo, per $y=y_0+b$. Esiste perciò, in base al teorema degli zeri, un *unico* $\overline{y}\in(y_0-b,y_0+b)$ tale che $g(\overline{x},\overline{y})=0$. Si determina così per ogni $x\in(x_0-\delta,x_0+\delta)$ un unico valore $y=f(x)\in(y_0-b,y_0+b)$ tale che $g(x,f(x))=0$; chiaramente si ha $y_0=f(x_0)$. Per completare la prima parte del teorema occorre mostrare che f è continua in $(x_0-\delta,x_0+\delta)$. A tale scopo sia $\varepsilon, 0<\varepsilon<b$. Se ripetiamo il ragionamento appena fatto con ϵ al posto di b troviamo un intorno $(x_0-\delta_\varepsilon,x_0+\delta_\varepsilon)$ ed una funzione f^* tale che $y_0=f^*(x_0), g(x,f^*(x))=0$ in $(x_0-\delta_\varepsilon,x_o+\delta_\varepsilon)$ e che $y_0-\varepsilon<f^*(x)<y_0+\epsilon$ nello stesso insieme.

Poiché, per unicità, $f^*=f$ in $(x_0-\delta_\varepsilon,x_0+\delta_\varepsilon)$, ne segue la continuità di f in x_0. Per gli altri punti $\overline{x}\in(x_0-\delta,x_0+\delta)$ è sufficiente osservare che $g_y(\overline{x},y(\overline{x}))>0$; si può quindi ragionare per $(\overline{x},y(\overline{x}))$ come è stato fatto per (x_0,y_0).

Per quanto riguarda la seconda parte del teorema, osserviamo che, per la formula del valor medio, si può scrivere, per ogni $(x,y)\in W, (\overline{x},\overline{y})\in W$:

$$g(x,y)-g(\overline{x},\overline{y})=g_x(\xi,\eta)(x-\overline{x})+g_y(\xi,\eta)(y-\overline{y}), \tag{3.6}$$

dove (ξ,η) è un punto opportuno sul segmento congiungente (x,y) con $(\overline{x},\overline{y})$, in particolare $(\xi,\eta)\in W$.

Se ora $\overline{y}=f(\overline{x})$ e $y=f(x)$ abbiamo $g(x,f(x))=g(\overline{x},f(\overline{x}))=0$; inoltre

$$g_y(\xi,\eta)\geq\min_W g_y=m>0\ .$$

Dalla (3.6) si ottiene allora

$$\frac{f(x)-f(\overline{x})}{x-\overline{x}}=-\frac{g_x(\xi,\eta)}{g_y(\xi,\eta)}\ . \tag{3.7}$$

Facendo tendere x a $\overline{x}$, per la continuità di f, $f(x)\to f(\overline{x})$, che implica $(\xi,\eta)\to(\overline{x},\overline{y})$.

Usando la continuità di g_x e g_y, dalla (3.7), passando al limite per $x \to \overline{x}$, si deduce

$$f'(\overline{x}) = -\frac{g_x(\overline{x}, \overline{y})}{g_y(\overline{x}, \overline{y})}$$

che è la (3.4). Da questa stessa formula e dal teorema di continuità delle funzioni composte segue che f' è continua in $(x_0 - \delta, x_0 + \delta)$. □

Nel teorema 3.8, presentato più avanti, rientra come caso particolare il teorema 3.1 e perciò avremmo potuto evitare di presentare la dimostrazione di quest'ultimo; l'abbiamo inserita ugualmente poiché fa uso di una tecnica elementare che riteniamo istruttiva.

La formula (3.4) permette di scrivere l'equazione della retta tangente al grafico della funzione implicita nel punto (x_0, y_0). Si ha:

$$y = y_0 - \frac{g_x(x_0, y_0)}{g_y(x_0, y_0)}(x - x_0) .$$

Disponendo di ulteriore regolarità per g si possono calcolare derivate di f di ordine superiore nel punto x_0. Se, in aggiunta alle ipotesi del teorema 3.1, supponiamo che $g \in C^2(A)$, dalla (3.4) e dal teorema di derivazione delle funzioni composte, deduciamo che, essendo g_x, g_y ed f di classe C^1, anche f' è di classe C^1; una formula per f'' si ottiene derivando il secondo membro della (3.4):

$$f'' = -\frac{[g_{xx} + g_{xy}f']g_y - g_x[g_{yx} + g_{yy}f']}{(g_y)^2}$$

e, tenendo conto che $f' = -g_x/g_y$ e $g_{xy} = g_{yx}$:

$$\begin{aligned} f'' &= -\frac{g_{xx}(g_y)^2 - 2g_{xy}g_xg_y + g_{yy}(g_x)^2}{(g_y)^3} = \\ &= -\frac{g_{xx} + 2g_{xy}y' + g_{yy}(y')^2}{g_y} . \end{aligned} \tag{3.8}$$

Con identico ragionamento si deduce che se $g \in C^k(A)$, allora $f \in C^k(U)$, se $g \in C^\infty(A)$ allora $f \in C^\infty(U)$. Si possono poi calcolare le derivate successive, per ottenere approssimazioni di Taylor dell'ordine desiderato, così come è stato indicato per f''.

Da un punto di vista pratico a volte conviene procedere derivando direttamente l'equazione.

Infatti, se pensiamo $y = f(x)$, l'equazione $g(x, y) = 0$ è una identità per $x \in U$; si può allora derivare rispetto ad x ottenendo

$$g_x + g_y y' = 0$$

da cui la formula (3.4). Derivando ulteriormente:

$$g_{xx} + 2g_{xy}y' + g_{yy}(y')^2 + g_y y'' = 0$$

e tenendo conto che $y' = -g_x/g_y$ si ritrova la formula (3.8) per la derivata seconda di f. Un'altra derivazione permetterebbe di ottenere una formula per y''' e così via per le derivate di ordine superiore.

***Esempio* 3.1** - Sia $g(x,y) = xe^y + ye^x = 0$. Tutte le ipotesi del teorema sono soddisfatte se $(x_0, y_0) = (0,0)$. Infatti $g \in C^\infty(\mathbb{R}^2), g(0,0) = 0, g_y(0,0) = 1$. Ne segue che in un intorno di $x_0 = 0, g(x,y) = 0$ definisce una funzione $y = f(x)$ di classe C^∞. Dalla (3.4), $f'(x) = -\dfrac{e^y + ye^x}{xe^y + e^x}$ e quindi $f'(0) = -1$. Essendo $f(0) = 0$ la tangente al grafico di f in $(0,0)$ è $y = -x$.

Alternativamente, si può derivare l'equazione rispetto ad x; si ha:

$$(xy' + 1)e^y + (y' + y)e^x = 0$$

e, ponendo $x = 0, y = 0$, si riottiene $y'(0) = -1$.
Derivando un'altra volta abbiamo:

$$(xy'^2 + 2y' + xy'')e^y + (y + 2y' + y'')e^x = 0 \ ,$$

da cui, sostituendo $x = 0, y = 0, y' = -1$, si deduce $y''(0) = 4$.
Calcoliamo anche la derivata terza mediante un'altra derivazione:

$$(xy'^3 + 3y'^2 + 3xy'y'' + 3y'' + xy''')e^y + (y + 3y' + 3y'' + y''')e^x = 0 \ ;$$

sostituendo $x = 0, y = 0, y' = -1$ e $y'' = 4$ si ricava $y'''(0) = -24$.
Possiamo ora scrivere la formula di Mac Laurin arrestata al 3° ordine per la funzione implicita $y = f(x)$:

$$f(x) = -x + 2x^2 - 4x^3 + o(x^3) \qquad \text{per } x \to 0 \ .$$

I ruoli di x e y nel teorema 3.1 sono intercambiabili. Se $g_y(x_0, y_0) = 0$ ma $g_x(x_0, y_0) \neq 0$ esiste un intorno V di y_0 in cui è definita implicitamente da $g(x,y) = 0$ una funzione $x = h(y)$; la formula (3.4) diventa

$$h'(y) = -\frac{g_y(h(y), y)}{g_x(h(y), y)} \ .$$

Restano esclusi dalle precedenti considerazioni i punti in cui $\nabla g = \mathbf{0}$. Questi punti si chiamano *singolari* o *critici*. L'analisi in un intorno di un punto singolare isolato, almeno nei suoi aspetti più elementari, è contenuta nel prossimo paragrafo. Essa fa intervenire in modo naturale la funzione $z = g(x,y)$; se (x_0, y_0) è un punto singolare in cui g è differenziabile, il piano tangente al grafico di $z = g(x,y)$

nel punto $(x_0, y_0, g(x_0, y_0))$ è orizzontale (essendo $\nabla g(x_0, y_0) = 0$) e, essendo $g(x_0, y_0) = 0$, coincide col piano $z = 0$.

Si osservi che la condizione $\nabla g(x_0, y_0) \neq \mathbf{0}$ si presenta geometricamente come una *condizione di trasversalità* di tale piano tangente rispetto al piano $z = 0$.

3.3 Insiemi di livello. Punti singolari

L'insieme E_0 dei punti del piano in cui $g(x, y) = 0$ può essere interpretato geometricamente come l'intersezione del grafico di $z = g(x, y)$ con il piano $z = 0$. Più in generale, per $c \in \mathbb{R}$, possiamo introdurre l'insieme E_c, proiezione sul piano $z = 0$ dell'insieme ottenuto intersecando il grafico di $z = g(x, y)$ con il piano $z = c$. Tale insieme si chiama *insieme di livello* di g.

Tutto ciò che è stato detto per l'equazione $g(x, y) = 0$ si trasferisce naturalmente all'equazione $g(x, y) = c$, poiché g e $g - c$ hanno le stesse proprietà.

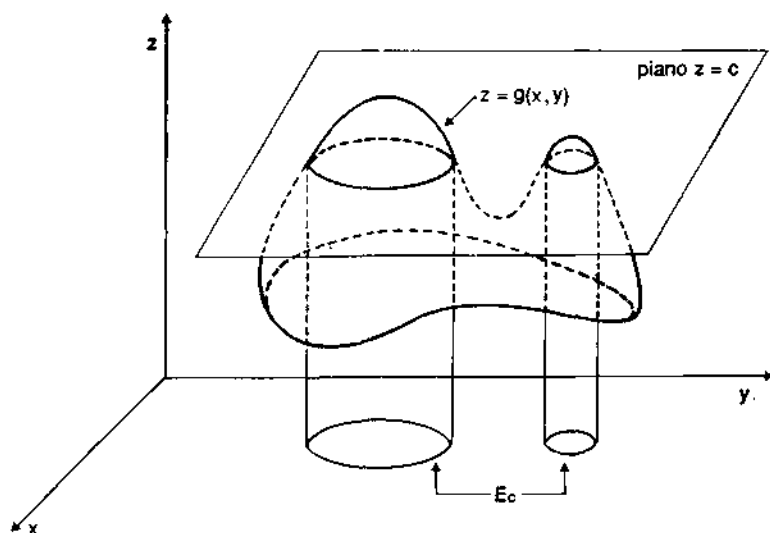



Fig. 7.18

Senza ipotesi su g, E_c può essere un qualunque insieme del piano, "selvaggio" fin che si vuole. Infatti se $A \subseteq \mathbb{R}^2$ e g è la funzione caratteristica di A, allora $E_1 = A$.

Dal teorema 3.1, se g è di classe C^1 nell'intorno di un punto (x_0, y_0) tale che $g(x_0, y_0) = c$ e $\nabla g(x_0, y_0) \neq \mathbf{0}$ allora E_c è rappresentabile localmente come grafico di una funzione. Se quindi $\nabla g \neq \mathbf{0}$ in ogni punto di E_c, tranne che eventualmente in un numero finito di punti isolati, è ragionevole chiamare E_c *curva* o *linea* piana; in tal caso parleremo di curve o *linee di livello* di g.

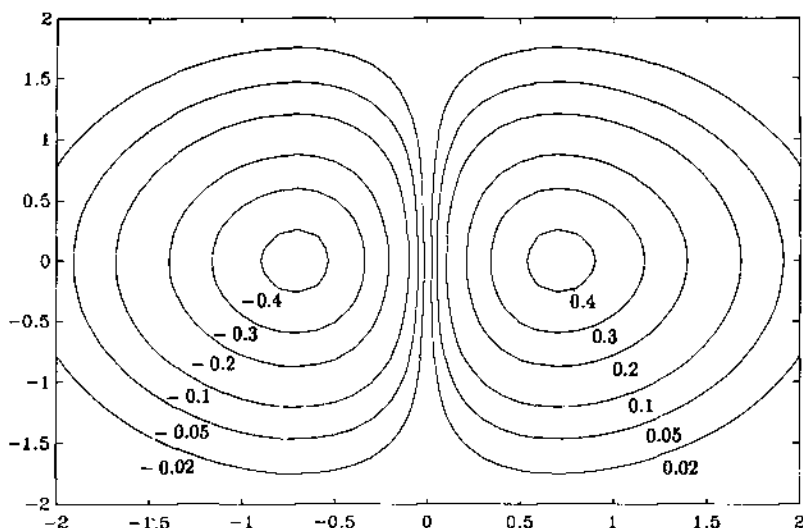



Fig. 7.19 Linee di livello di $z = x \cdot \exp(x^2 - y^2)$.

Come conseguenza della continuità di g, E_c risulta un insieme chiuso (in $\mathbb{R}^2$) per ogni $c \in \mathbb{R}$; in generale non sarà connesso ma unione di curve connesse, anche infinite come ad esempio nel caso

$$\sin(x - y) = 0$$

in cui E_0 è unione delle rette $x - y = k\pi$ con $k \in \mathbb{Z}$.

Ognuna delle curve connesse che compongono E_0 si chiama *ramo* della curva; il termine ramo si usa anche per indicare un tratto di curva rappresentabile localmente come grafico di una funzione. Ad esempio i grafici di $y = \sqrt{1 - x^2}$ e $x = -\sqrt{1 - y^2}$ sono *rami* della circonferenza $x^2 + y^2 = 1$.

Se $g(x, y)$ è un polinomio di grado m in x e y, la curva definita da $g(x, y) = 0$ si dice *algebrica* di grado m.

I più noti esempi studiati fin dall'antichità di curve algebriche sono le (sezioni) coniche (*): ellissi, parabole, iperboli. Sono curve di secondo grado la cui equazione generale è

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0 .$$

(*) Come scopritore delle coniche si ritiene generalmente *Menecmo* ($\sim$ 350 a.C.) discepolo di Platone ed Eudosso.

Mediante un'opportuna rototraslazione, nei casi non degeneri, si può ricondurre l'equazione ad una delle forme canoniche seguenti:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1 \qquad \text{(Ellisse di semiassi } a \text{ e } b\text{)} .$$

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1 \qquad \text{(Iperbole di asintoti } y = \pm\frac{b}{a}x\text{)}$$

$$y = ax^2 \qquad \text{(Parabola con fuoco } F\left(0, \frac{1}{4a}\right)\text{)} .$$

Altre curve algebriche sono le seguenti, che studieremo in qualche dettaglio più avanti

$$(x^2 + y^2)^2 - 2a^2(x^2 - y^2) = 0 \qquad \text{(Lemniscata di Bernoulli)}$$

$$x^3 + y^3 - 3axy = 0 \qquad \text{(Folium di Cartesio)}$$

$$x(x^2 + y^2) - 2ry^2 = 0 \qquad \text{(Cissoide di Diocle)} .$$

Ritornando alle curve di livello E_c di una funzione g di classe C^1, osserviamo che la formula (3.4), che assegna il coefficiente angolare della retta tangente ad E_c in un punto *non* singolare, ha una notevole conseguenza.

Scrivendo l'equazione della tangente nella forma

$$(x - x_0)g_x(x_0, y_0) + (y - y_0)g_y(x_0, y_0) = 0 \tag{3.9}$$

si vede che il vettore $\nabla g(x_0, y_0)$ è ortogonale ad essa. Infatti la (3.9) esprime l'annullarsi del prodotto scalare tra il vettore $\nabla g(x_0, y_0)$ e il vettore $(x - x_0, y - y_0)$, diretto lungo la tangente (cfr. fig. 7.20).

Poiché abbiamo visto che la direzione di ∇g è quella di massima crescita per g otteniamo la conclusione seguente.

Proposizione 3.2 - *Sia $g \in C^1(A)$, A aperto di $\mathbb{R}^2$. Se E_c è privo di punti singolari, in ogni punto di E_c la tangente è ortogonale alla direzione di massima crescita di g in quel punto.*

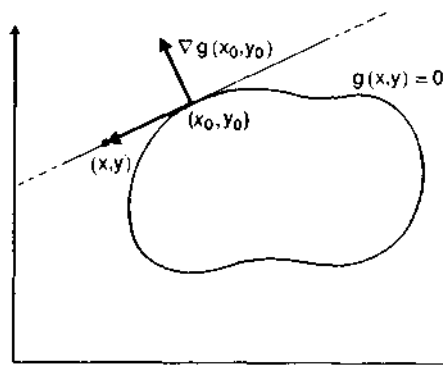



Fig. 7.20

La famiglia delle curve di livello interviene in svariate situazioni concrete. Ad esempio i topografi usano le linee di livello altimetriche (le cosiddette *isoipse* o curve di quota costante) per descrivere l'andamento dei pendii. In una regione (piana) sede di un campo di temperatura, le linee di livello costituiscono le *isoterme* o curve i cui punti sono alla stessa temperatura. La proposizione 3.2 indica che il *flusso di calore* che segue le linee di massima crescita della temperatura, è normale alla *isoterme*. Si dice allora che linee di flusso e linee di livello costituiscono una famiglia di *traiettorie ortogonali.*

Veniamo ora all'analisi del comportamento di una curva $g(x, y) = 0$ nell'intorno di un suo punto singolare (x_0, y_0). Assumeremo che nell'intorno in questione g sia di classe C^2 e, per il momento, che in (x_0, y_0) *non* tutte le derivate seconde di g siano nulle.

Una prima idea è quella di studiare le intersezioni del fascio di rette di centro in (x_0, y_0) con la curva, in un intorno di (x_0, y_0). Scriviamo la generica retta del fascio nella forma

$$x = x_0 + at, y = y_0 + bt \tag{3.10}$$

dove a e b sono i coseni direttori: $a^2 + b^2 = 1$.

Usando la formula di Taylor arrestata al secondo ordine, con centro in (x_0, y_0) possiamo poi scrivere l'equazione $g(x, y) = 0$ nel modo seguente, ricordando che $\nabla g(x_0, y_0) = \mathbf{0}$:

$$\begin{aligned} g_{xx}(x_0, y_0)(x - x_0)^2 + 2(x - x_0)(y - y_0)g_{xy}(x_0, y_0) + \\ + g_{yy}(x_0, y_0)(y - y_0)^2 + o(\rho^2) = 0 \end{aligned}$$

per $\rho \to 0$, dove $\rho = \sqrt{(x - x_0)^2 + (y - y_0)^2}$.

Sostituendo in questa equazione le (3.10) otteniamo:

$$t^2\{a^2 g_{xx}(x_0, y_0) + 2ab g_{xy}(x_0, y_0) + b^2 g_{yy}(x_0, y_0) + o(1)\} = 0 \tag{3.11}$$

per $t \to 0$.

Una prima soluzione è $t = 0$ che corrisponde al punto (x_0, y_0).

È importante notare che questa è una soluzione *doppia* della (3.11), comparendo al primo membro il fattore t^2. Per questa ragione il punto (x_0, y_0) si chiama anche *punto doppio.*

Rimuovendo il fattore t^2 dalla (3.11), l'equazione rimanente è, per $t \to 0$,

$$a^2 g_{xx}(x_0, y_0) + 2ab g_{xy}(x_0, y_0) + b^2 g_{yy}(x_0, y_0) + o(1) = 0\ . \tag{3.12}$$

Ci chiediamo ora se è possibile che la retta (3.10) intersechi la curva in qualche altro punto (x, y) che tenda ad (x_0, y_0) quando la retta stessa tenda ad assumere una particolare posizione limite "tangente", ad esempio come illustrato nella figura 7.21.

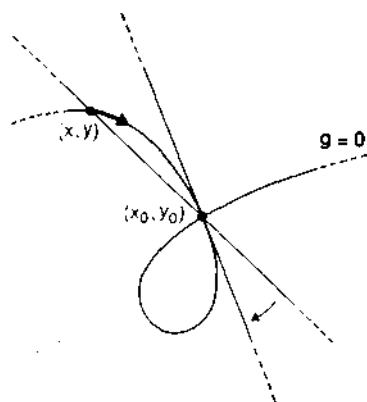



Fig. 7.21

A questo scopo osserviamo che se $(x, y) \to (x_0, y_0)$ anche $t \to 0$ e quindi $o(1) \to 0$. Dalla (3.12) si deduce allora che per la posizione limite della retta si deve avere:

$$a^2 g_{xx}(x_0, y_0) + 2ab g_{xy}(x_0, y_0) + b^2 g_{yy}(x_0, y_0) = 0 \ . \tag{3.13}$$

La (3.13) è un'equazione di secondo grado per il rapporto a/b o b/a che fissa la direzione della retta limite.

La discussione dipende ora dal segno dell'espressione

$$H(x_0, y_0) = g_{xx}(x_0, y_0) g_{yy}(x_0, y_0) - g_{xy}(x_0, y_0)^2$$

che, peraltro, coincide con il determinante di $\mathbf{H}_g(x_0, y_0)$, la matrice Hessiana di g nel punto (x_0, y_0).

Distinguiamo 3 casi

Caso 1: $H(x_0, y_0) > 0$.

La (3.13) non ha soluzioni reali, dunque non esistono tangenti (reali). Il punto (x_0, y_0) si dice allora punto *singolare isolato* della curva $g = 0$.

È il caso dell'equazione, già considerata nel paragrafo 1,

$$x^2 + y^2 = 0 \qquad \text{oppure dell'equazione} \qquad (x^2 - 1)^2 + (y^2 - 1)^2 - 2 = 0$$

dove $(x_0, y_0) = (0, 0)$.

La situazione nel caso 1 può essere interpretata facilmente in termini di convessità (concavità) di g nell'intorno di (x_0, y_0). Osserviamo infatti che $g_{xx}(x_0, y_0)$ e $g_{yy}(x_0, y_0)$ non sono entrambe nulle; per fissare le idee sia $g_{yy}(x_0, y_0) > 0$. Poiché g è di classe C^2, per il teorema della permanenza del segno, esiste un intorno di (x_0, y_0), diciamo W, in cui $H > 0$ e $g_{yy} > 0$.

In tale intorno, allora, g è strettamente convessa e perciò il suo grafico ha come unica intersezione col piano tangente nel punto $(x_0, y_0, 0)$, ovvero il piano $z = 0$, il punto stesso, che risulta di conseguenza un punto singolare isolato della linea di livello E_0. Si dice anche che (x_0, y_0) è un *punto ellittico* per $z = g(x, y)$.

Caso 2: $H(x_0, y_0) < 0$.

La (3.13) individua due rette reali e distinte. In effetti queste due rette sono tangenti a due *rami* della curva che si incrociano nel punto (x_0, y_0). Più precisamente si può dimostrare la seguente proposizione:

Proposizione 3.3 - *Se* $H(x_0, y_0) < 0$, *in un intorno di* (x_0, y_0) *la curva* $g(x, y) = 0$ *è unione dei grafici di due funzioni* $y = f_1(x)$ *e* $y = f_2(x)$ *di classe* C^1 *(in un intorno di* x_0*), intersecantisi in* (x_0, y_0) *ed aventi come tangenti in questo punto quelle individuate dalla* (3.13).

In questo caso (x_0, y_0) si chiama *punto doppio ordinario* o *nodo.*

Esempi

3.2. Il prototipo di nodo è l'origine per la conica degenere $x^2 - y^2 = 0$ che si riduce alle due rette $y = \pm x$.

3.3. La Lemniscata di Bernoulli,

$$(x^2 + y^2)^2 - 2a^2(x^2 - y^2) = 0 \qquad (a \neq 0)$$

ha nell'origine ($H(0,0) = -1$) un nodo con tangenti $y = \pm x$. Per $a = 2\sqrt{2}$ la curva è disegnata in figura 7.22.

3.4. Il Folium di Cartesio,

$$x^3 + y^3 - 3axy \qquad (a \neq 0)$$

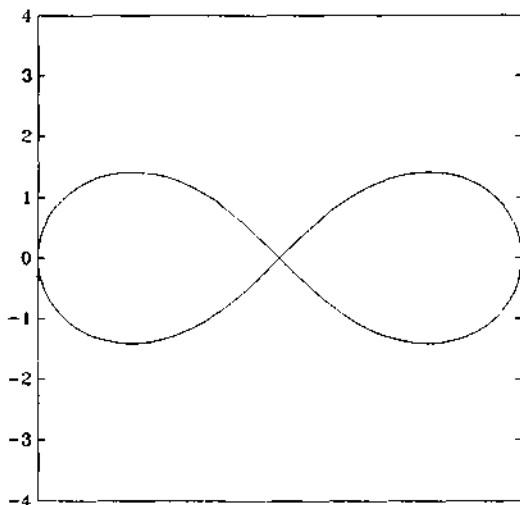



Fig. 7.22 Lemniscata di Bernoulli.

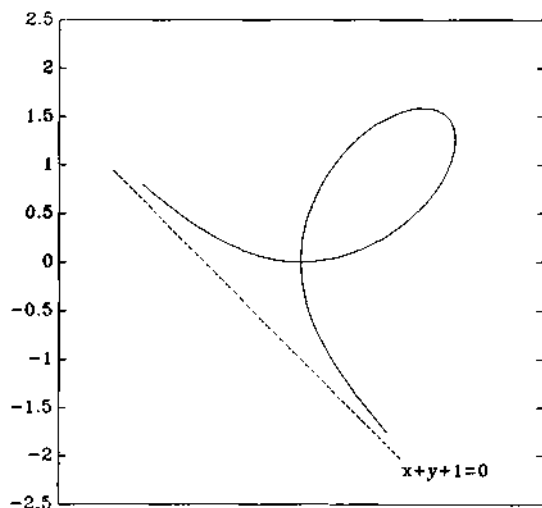



Fig. 7.23 Folium di Cartesio.

ha nell'origine ($H(0,0) = -9a^2$) un nodo con tangenti coincidenti con gli assi coordinati. Per $a = 1$ la curva è disegnata in figura 7.23

In relazione al grafico di $z = g(x, y)$ il punto (x_0, y_0) si chiama *punto iperbolico*: il piano tangente $z = 0$ attraversa il grafico e lo interseca nell'intorno di (x_0, y_0) lungo le due curve

$$y = f_1(x) \quad \text{e} \quad y = f_2(x) \ .$$

Dimostrazione della proposizione 3.3 - In base ai calcoli nell'esempio 2.11, con una trasformazione ortogonale di coordinate ci si può sempre ricondurre al caso in cui $g_{yy}(x_0, y_0) \neq 0$; sia $g_{yy}(x_0, y_0) > 0$ (il caso opposto è analogo). Facciamo vedere che esistono le due funzioni $y = f_1(x)$ e $y = f_2(x)$. Osserviamo innanzitutto che, essendo g_{yy} continua, per il teorema della permanenza del segno, esiste un rettangolo $[x_0 - \delta, x_0 + \delta] \times [y_0 - b, y_0 + b]$ in cui $g_{yy} > 0$ e $H(x, y) < 0$. Ciò implica subito che le sezioni $z = g(\overline{x}, y)$ con $\overline{x} \in [x_0 - \delta, x_0 + \delta]$ sono *strettamente convesse* in $[y_0 - b, y_0 + b]$. La strategia consisterà ora nel dimostrare che per x vicino ad x_0, $x \neq x_0$, queste sezioni hanno un minimo negativo in un punto $y = \varphi(x)$, mentre agli estremi $y = y_0 - b$, $y = y_0 + b$ assumono valore positivo. Dividiamo la dimostrazione in 3 passi.

a) Consideriamo la sezione $z = g(x_0, y)$; poiché $g(x_0, y_0) = g_y(x_0, y_0) = 0$ tale sezione assume valore minimo (zero) in $y = y_0$ ed è positiva in $y = y_0 + b$ e $y = y_0 - b$. Per il teorema della permanenza del segno, esiste $\eta \leq \delta$ tale che $g(x, y_0 - b) > 0$ e $g(x, y_0 + b) > 0$ per ogni $x \in (x_0 - \eta, x_0 + \eta)$.

b) Consideriamo ora una sezione $z = g(\overline{x}, y)$ per $\overline{x} \in (x_0 - \eta, x_0 + \eta)$. L'equazione $g_y(\overline{x}, y) = 0$, essendo $g_{yy}(\overline{x}, y) > 0$, per il teorema 3.1 definisce implicitamente $y = \varphi(\overline{x})$ per ogni $\overline{x}$ in un intorno di x_0 che possiamo chiamare ancora $(x_0 - \eta, x_0 + \eta)$. Il punto $y = \varphi(\overline{x})$ è dunque un punto stazionario per $z = g(\overline{x}, y)$, strettamente convessa e pertanto

risulta un punto di minimo. Sempre in conseguenza del teorema 3.1 risulta che, se $x \in (x_0 - \eta, x_0 + \eta)$, $\varphi(x) \in [y_0 - b, y_0 + b]$, che φ è di classe C^1 e che vale la formula:

$$\varphi'(x) = -\frac{g_{yx}(x, \varphi(x))}{g_{yy}(x, \varphi(x))} . \tag{3.14}$$

Chiamiamo $y = \varphi(x)$ *linea dei minimi*. Osserviamo che $y_0 = \varphi(x_0)$ e che $H(x, \varphi(x)) < 0$ in $(x_0 - \eta, x_0 + \eta)$.

c) Studiamo ora la funzione $\Psi(x) = g(x, \varphi(x))$ (= quota assunta da g sulla linea dei minimi). Abbiamo $\Psi(x_0) = g(x_0, y_0) = 0$; inoltre:

$$\Psi'(x) = g_x(x, \varphi(x)) + g_y(x, \varphi(x))\varphi'(x) = g_x(x, \varphi(x))$$

essendo $g_y(x, \varphi(x)) = 0$, e quindi $\Psi'(x_0) = g_x(x_0, y_0) = 0$; usando poi la (3.14) e ricordando che $H(x, \varphi(x)) < 0$:

$$\Psi''(x) = g_{xx}(x, \varphi(x)) + g_{yx}(x, \varphi(x))\varphi'(x) = H(x, \varphi(x))/g_{yy}(x, \varphi(x)) < 0 .$$

Concludiamo che Ψ è concava strettamente ed ha un massimo in $x = x_0$ dove vale zero. Dunque $\Psi(x) < 0$ se $x \neq x_0$. La situazione è quella illustrata in figura 7.24.

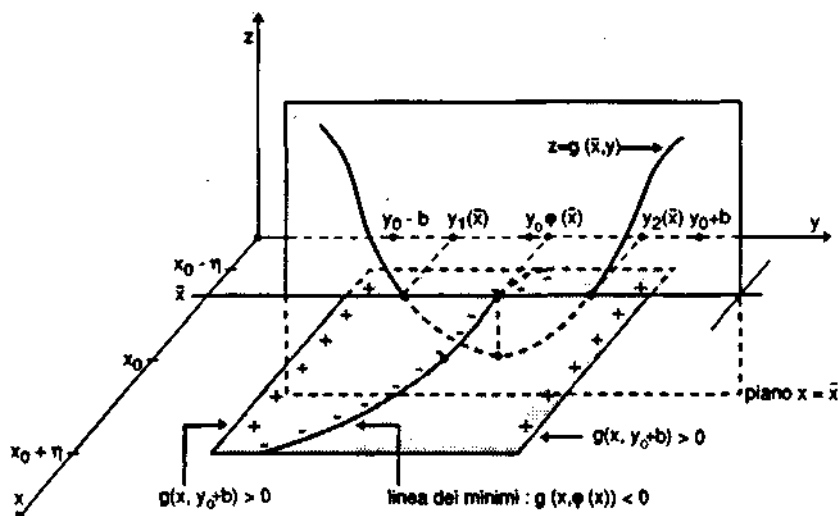



Fig. 7.24

Si vede che, per ogni $\bar{x} \in (x_0 - \eta, x_0 + \eta)$ esistono esattamente due zeri di $g(\bar{x}, y)$ in corrispondenza ai punti $y_1(\bar{x}), y_2(\bar{x})$ con $y_1(\bar{x}) < \varphi(\bar{x}) < y_2(\bar{x})$. Al variare di x in $(x_0 - \eta, x_0 + \eta)$ i punti $y_1(x)$ e $y_2(x)$ descrivono i due rami richiesti. La costruzione stessa (come nel teorema 3.1) indica che y_1 ed y_2 sono continue. L'ultima parte della dimo-

strazione consiste nel rendere più rigoroso il ragionamento che ha condotto alla (3.13). Omettiamo i dettagli (si veda comunque l'esercizio 17.). □

Caso $H(x_0, y_0) = 0$.

La (3.13) individua due rette coincidenti. Per la corretta analisi di questo caso occorre far intervenire le derivate di g di ordine superiore al secondo (*). Ci limitiamo ad indicare con esempi le possibili situazioni.

In questo caso il punto (x_0, y_0) si dice *parabolico* per la funzione $z = g(x, y)$.

Esempi

3.5. Sia $g(x, y) = (x + y)^2 = 0$. Allora $H(0,0) = 0$ e la curva stessa si riduce a due rette sovrapposte di equazione $y = -x$.

3.6. Sia $g(x, y) = x(x^2 + y^2) - 2ry^2 = 0 \qquad r \neq 0$.

Per questa curva algebrica, che si chiama *Cissoide di Diocle*, abbiamo $H(0,0) = = 0$. L'unica tangente (contata due volte) nell'origine è $y = 0$.

In questo caso l'origine si chiama *cuspide*. Per $r = 1$ la curva è disegnata in figura 7.25.

Altri tipi di cuspidi si presentano nei casi seguenti

$$g(x, y) = y^2 - x^4 = (y - x^2)(y + x^2)$$

$$g(x, y) = y^2 - 3x^2y + 2x^4 = (y - x^2)(y - 2x^2) .$$

Se tutte le derivate di g del secondo ordine sono uguali a zero in (x_0, y_0), ovvero se $d^2g(x_0, y_0) = 0$, l'analisi diventa molto più complicata. Ci limitiamo ad osservare che in (x_0, y_0) possono incrociarsi più rami della curva dando origine a singolarità di vario tipo.

Esempi

3.7. La curva algebrica di equazione

$$(x^2 + y^2)^2 - y(y^2 - x^2) = 0$$

ha nell'origine un punto triplo. Intersecando con il fascio di rette con centro in $(0,0)$, con un ragionamento simile a quello che ha portato alla (3.13) si ricava che

(*) L'analisi procede come nel caso $H(x_0, y_0) < 0$ fino al punto c) nella dimostrazione della proposizione 3.3. A quel punto si otterrebbe $\Psi'(x_0) = \Psi''(x_0) = 0$. Per avere informazioni sull'andamento di Ψ occorre quindi calcolare, ad esempio, la prima derivata di Ψ che non si annulla in x_0.

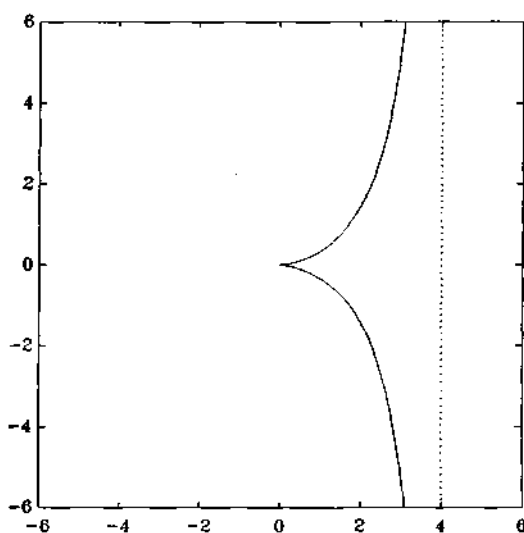



Fig. 7.25 Cissoide di Diocle.

le tangenti ai rami della curva si ottengono uguagliando a zero l'insieme dei termini di grado inferiore: $y(x^2 - y^2) = 0$; da cui $y = 0$ e $y = \pm x$. L'origine si chiama *doppio nodo* (fig. 7.26).

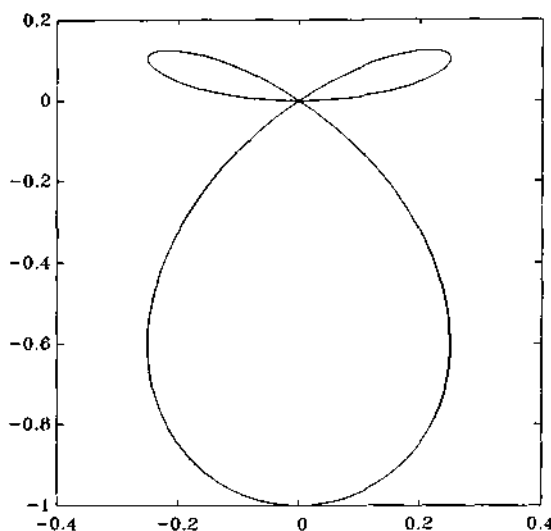



Fig. 7.26 La curva algebrica $(x^2 + y^2)^2 - y(y^2 - x^2) = 0$.

3.8. La curva algebrica di equazione

$$(x^2 + y^2)^2 - 3xy^2 = 0$$

ha nell'origine un punto triplo. Le tangenti sono $x = 0$ e $y = 0$ (contata 2 volte). La curva è disegnata in figura 7.27.

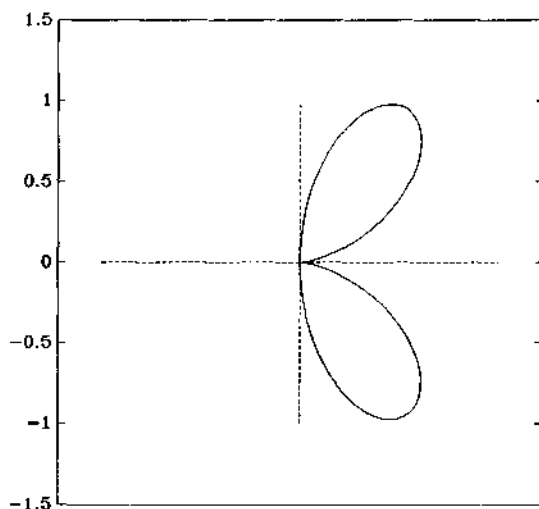



Fig. 7.27 La curva algebrica $(x^2 + y^2)^2 - 3xy^2 = 0$.

3.4 Inviluppo di una famiglia di curve

Siano A un aperto di $\mathbb{R}^2$ e $I \subseteq \mathbb{R}$ un intervallo. Data $\Psi = \Psi(x, y, c), (x, y) \in A$ e $c \in I$, di classe C^1 in $A \times I$, l'equazione

$$\Psi(x, y, c) = 0 \tag{3.15}$$

assegna una famiglia di curve ad un parametro.

Definizione 3.2 - *Si dice inviluppo della famiglia* (3.15) *una curva tangente, in ogni suo punto non singolare, ad una ed una sola curva della famiglia.*

La nozione di inviluppo precisa quella intuitiva di "luogo dei punti di intersezione tra curve infinitamente vicine". Se assumiamo che questo "luogo" esista, nei suoi punti saranno soddisfatte le equazioni $\Psi(x, y, c) = 0$ e $\Psi(x, y, c+h) = 0$ per h

"abbastanza piccolo", e, di conseguenza, anche

$$\frac{\Psi(x,y,c+h)-\Psi(x,y,c)}{h}=0 .$$

Passando al limite per $h \to 0$ si ottiene che i punti del "luogo" devono soddisfare anche l'equazione $\Psi_c(x,y,c)=0$.

Queste considerazioni euristiche conducono alla seguente proposizione:

Proposizione 3.4 - *Se esiste l'inviluppo della famiglia ed è localmente rappresentato da due equazioni parametriche $x = x(c)$ e $y = y(c)$ di classe $C^1(I'), I' \subseteq I$, allora per ogni $c \in I'$, sono soddisfatte le equazioni*;

$$\Psi(x(c),y(c),c)=0, \Psi_c(x(c),y(c),c)=0 . \qquad (3.16)$$

Dimostrazione - Se $x = x(c)$ e $y = y(c)$ sono le equazioni dell'inviluppo allora la condizione di tangenza nel punto $x(c), y(c)$ si scrive:

$$\Psi_x(x(c),y(c),c)x'(c)+\Psi_y(x(c),y(c),c)y'(c)=0 \qquad (3.17)$$

essendo $(x'(c), y'(c))$ un vettore tangente all'inviluppo e $\nabla\Psi(x(c),y(c),c)$ normale alla curva $\Psi(x,y,c)=0$.

Poiché ogni punto $(x(c),y(c))$ dell'inviluppo appartiene alla curva $\Psi(x,y,c)=0$, abbiamo, per ogni $c \in I'$:

$$\Psi(x(c),y(c),c)=0$$

dalla quale, derivando rispetto a c:

$$\Psi_x(x(c),y(c),c)x'(c)+\Psi_y(x(c),y(c),c)y'(c)+\Psi_c(x(c),y(c),c)=0 \qquad (3.18)$$

ed infine, ricordando la (3.17) si ottiene $\Psi_c(x(c),y(c),c)=0$. □

Dal punto di vista pratico, la proposizione conduce alla seguente regola: per ottenere l'inviluppo di una famiglia di curve $\Psi(x,y,c)=0$, si considera il sistema

$$\begin{cases} \Psi(x,y,c)=0 \\ \Psi_c(x,y,c)=0 \end{cases} \qquad (3.19)$$

e si cerca di ricavare x e y in funzione di c (*). Alternativamente si ricava $c = c(x,y)$ dalla seconda e si scrive l'inviluppo nella forma $\Psi(x,y,c(x,y))=0$ (**).

(*) Ciò si può certamente fare se $\frac{\partial(\Psi,\Psi_c)}{\partial(x,y)} \neq 0$, come risulterà dal teorema 3.8.

(**) Ciò si può certamente fare se $\Psi_{cc}(x,y,c) \neq 0$, come risulterà dal teorema 3.5.

Osservazione 3.1 - Le condizioni (3.16) sono solo necessarie per l'esistenza dell'inviluppo. Infatti se $x = x(c), y = y(c)$ è una coppia di equazioni parametriche che verifica le (3.16), dalla (3.18) otteniamo

$$\Psi_x(x(c), y(c), c)x'(c) + \Psi_y(x(c), y(c), c)y'(c) = 0 \ . \tag{3.20}$$

La (3.20) è verificata se $\Psi_x(x(c), y(c), c) = \Psi_y(x(c), y(c), c) = 0$ ovvero se, anziché l'inviluppo, $x = x(c)$ e $y = y(c)$ descrivono il *luogo dei punti singolari* della famiglia di curve.

Un'altra possibilità è che $x'(c) = y'(c) = 0$ e cioè che la linea $x = x(c), y = = y(c)$ non abbia una tangente ben definita (vedi esercizio 9.).

Dunque, una volta risolto il sistema (3.19) bisogna procedere all'identificazione di ciò che si è trovato.

Esempi

3.9. Sia $\Psi(x, y, c) = (x - c)^2 + y^2 - 1 = 0$ (famiglia di cerchi con centro $(c, 0)$ e raggio 1).

Risolvendo il sistema

$$\begin{cases} (x - c)^2 + y^2 - 1 = 0 \\ 2(x - c) = 0 \end{cases}$$

si trova $y^2 - 1 = 0$ che è l'unione delle due rette $y = \pm 1$ e che costituisce l'inviluppo della famiglia (fig. 7.28).

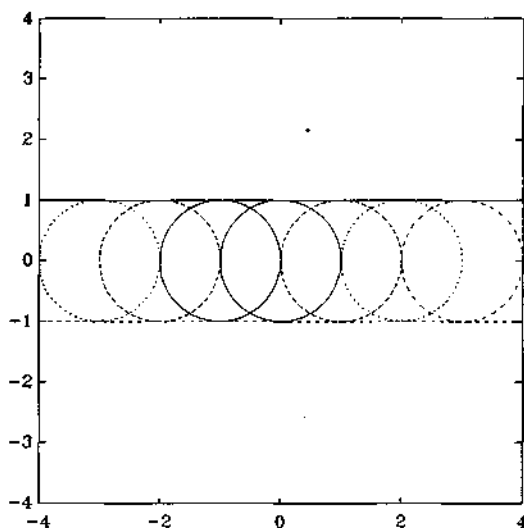



Fig. 7.28 Cerchi con inviluppo.

3.10. Sia $\Psi(x,y,c) = \dfrac{x^2}{c^2} + \dfrac{y^2}{(1-c)^2} - 1 = 0$ (famiglia di ellissi con centro $(0,0)$ e semiassi di lunghezza c e $1-c, 0 < c < 1$).

Risolvendo il sistema

$$\begin{cases} \dfrac{x^2}{c^2} + \dfrac{y^2}{(1-c)^2} - 1 = 0 \\ -\dfrac{2x^2}{c^3} + \dfrac{2y^2}{(1-c)^3} = 0 \end{cases}$$

si trova $y = (1-c)^{3/2}$ e $x = c^{3/2}$ ovvero

$$x^{\frac{2}{3}} + y^{\frac{2}{3}} = 1 \; .$$

Questa curva si chiama *astroide* (vedi fig. 7.29); si noti la presenza di cuspidi nei punti $(0, \pm 1)$ e $(\pm 1, 0)$.

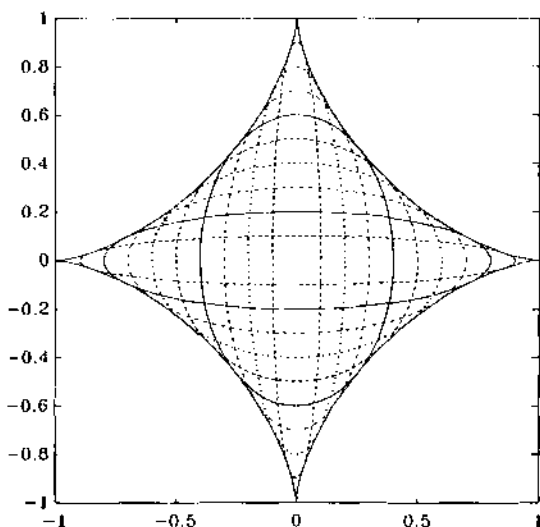



Fig. 7.29 Astroide come inviluppo di ellissi.

3.11. Sia $\Psi(x,y,c) = [y^2 + (x-c)^2](y-2) + y = 0$ (famiglia di *strofoidi*).

Risolvendo il sistema

$$\begin{cases} [y^2 + (x-c)^2](y-2) + y = 0 \\ 2(x-c)(y-2) = 0 \end{cases}$$

si ottiene $y^2(y-2)+y = y(y-1)^2 = 0$ ovvero le due rette $y = 0$ e $y = 1$ (contata due volte).

La retta $y = 0$ è l'inviluppo, $y = 1$ è il luogo dei punti singolari (nodi) della famiglia (cfr. fig. 7.30).

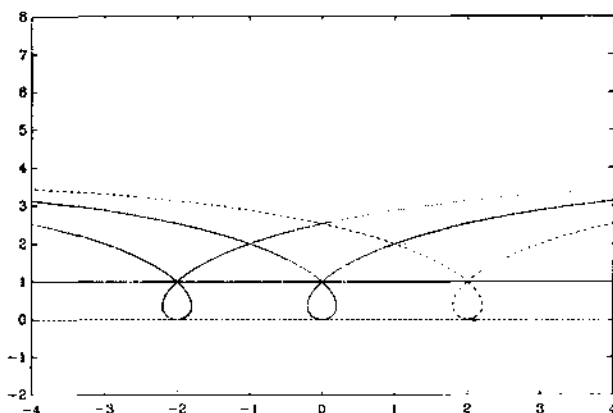



Fig. 7.30 Strofoidi con inviluppo e luogo di nodi.

3.12. Sia $\Psi(x,y,c) = (x-c)^2 - y^3 = 0$.
Risolvendo il sistema

$$\begin{cases} (x-c)^2 - y^3 = 0 \\ 2(x-c) = 0 \end{cases}$$

si trova $y = 0$, che non è inviluppo bensì luogo di cuspidi della famiglia (fig. 7.31).

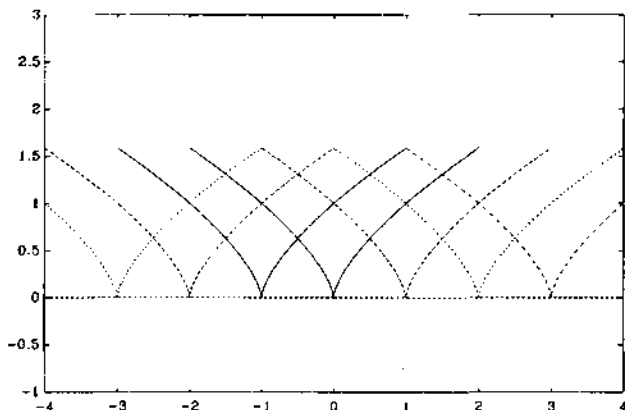



Fig. 7.31 Famiglia di cubiche con luogo di cuspidi.

3.5 Il teorema delle funzioni implicite in più di due variabili

Il teorema 3.1 può essere esteso al caso di funzioni di n variabili nel modo seguente. Per i punti $(x_1, x_2, \ldots, x_n, x_{n+1}) \in \mathbb{R}^{n+1}$, privilegiamo una variabile, ad esempio l'ultima, ponendo $\mathbf{x} = (x_1, x_2, \ldots, x_n), x_{n+1} = y$.

▪ **Teorema 3.5** - (di Dini). *Sia* $g : \mathbb{R}^{n+1} \supseteq A \to \mathbb{R}, A$ *aperto. Supponiamo che*

i) g *e* g_y *siano continue in* A

ii) nel punto $(\mathbf{x}^0, y_0) \in A$ *si abbia* $g(\mathbf{x}^0, y_0) = 0$ *e* $g_y(\mathbf{x}^0, y_0) \neq 0$.

Allora esistono un intorno U *di* $\mathbf{x}^0$ *ed un'unica funzione* $f : U \to \mathbb{R}$, *continua in* U, *tale che* $f(\mathbf{x}^0) = y_0$ *e che* $g(\mathbf{x}, f(\mathbf{x})) = 0$ *per ogni* $x \in U$.
Se inoltre

iii) $g_{x_1}, g_{x_2}, \ldots, g_{x_n}$ *sono continue in* A (dunque $g \in C^1(A)$) *allora* $f \in C^1(U)$ *e vale la formula*

$$f_{x_j}(\mathbf{x}) = -\frac{g_{x_j}(\mathbf{x}, f(\mathbf{x}))}{g_y(\mathbf{x}, f(\mathbf{x}))} \qquad \forall \mathbf{x} \in U, \forall j = 1, \ldots, n \,. \tag{3.21}$$

La funzione f si dice definita implicitamente dall'equazione $g(\mathbf{x}, y) = 0$.

Omettiamo la dimostrazione del teorema 3.5 poiché si può condurre sulla falsariga di quella del teorema 3.1, con ovvi cambiamenti dovuti alla dimensione superiore; inoltre, anche il teorema 3.5 è un caso particolare del teorema 3.8.
Il teorema 3.5 afferma che, se $g(x_1^0, x_2^0, \ldots, x_n^0, y_0) = 0$ e se $g_y(x_1^0, x_2^0, \ldots, x_n^0, y_0) \neq 0$ allora l'equazione

$$g(x_1, x_2, \ldots, x_n, y) = 0 \tag{3.22}$$

permette di esprimere y in funzione di $x_1, \ldots, x_n$ in un intorno di $\mathbf{x}^0$. Naturalmente il ruolo delle variabili è intercambiabile: ciascuna può prendere il posto di y.

Ancora, la regolarità di g si trasmette ad f : $g \in C^k(A) \Rightarrow f \in C^k(U), g \in C^\infty(A) \Rightarrow f \in C^\infty(U)$.

Per successive derivazioni della (3.22), pensando $y = f(x_1, \ldots, x_n)$, è poi possibile ricavare le derivate di ordine superiore di f in $(\mathbf{x}^0, y_0)$ allo scopo di approssimare f con polinomi di Taylor centrati in quel punto.

***Esempio* 3.13** - Sia $g(x, y, z) = \operatorname{arctg} z + xy^2 + xz - y^3 - 1 = 0$. La funzione g è di classe $C^\infty(\mathbb{R}^3)$. Nel punto $(0, -1, 0)$ si ha: $g(0, -1, 0) = 0$ e $g_z(0, -1, 0) = 1$. Dunque, in base al teorema 3.5 l'equazione $g(x, y, z) = 0$ definisce implicitamente $z = f(x, y)$ in un intorno di $(0, -1)$. Tale funzione risulta di classe C^∞.

Calcoliamo $df(0,-1)$ e $d^2f(0,-1)$. Si ha, dalla (3.21):

$$f_x(0,-1) = -\frac{g_x(0,-1,0)}{g_z(0,-1,0)} = -\frac{1}{1} = -1$$

$$f_y(0,-1) = -\frac{g_y(0,-1,0)}{g_z(0,-1,0)} = -\frac{-3}{1} = +3 .$$

Quindi $df(0,-1) = -dx + 3dy$.

Per calcolare il differenziale secondo differenziamo due volte l'equazione $g(x,y,z) = 0$, pensando $z = f(x,y)$; ricordando la proprietà di invarianza del differenziale primo, abbiamo:

$$dg = g_x dx + g_y dy + g_z dz = 0 .$$

Differenziando ancora abbiamo

$$d^2g + g_z d^2 z = 0 .$$

Poiché $d^2g(0,-1,0) = 6dy^2 - 4dxdy + 2dxdz$, ricaviamo, ricordando che $dz = df = -dx + 3dy$ e $g_z(0,-1,0) = 1$:

$$d^2 f = -6dy^2 + 4dxdy - 2dx(-dx + 3dy) = 2dx^2 - 2dxdy - 6dy^2 .$$

In conclusione possiamo scrivere, usando la formula di Taylor con centro in $(0,-1)$:

$$f(dx, -1 + dy) = -dx + 3dy + \frac{1}{2}(2dx^2 - 2dxdy - 6dy^2) + \\ + o\left(\sqrt{dx^2 + dy^2}\right) \quad \text{per } (dx, dy) \to (0,0) .$$

Equazioni della forma $g(x,y,z) = 0$, con g di classe C^1, nell'intorno dei punti in cui $\nabla g \neq \mathbf{0}$, rappresentano grafici di funzioni $z = z(x,y)$ (se $g_z \neq 0$) o $y = y(x,z)$ (se $g_y \neq 0$) o infine $x = x(y,z)$ (se $g_x \neq 0$).

In analogia col caso bidimensionale i punti in cui $\nabla g = \mathbf{0}$ si chiamano *punti singolari*.

La formula (3.21) permette di ricavare l'equazione del piano tangente in un punto *non* singolare (x_0, y_0, z_0) al grafico della funzione implicita. Se, ad esempio $g_z(x_0, y_0, z_0) \neq 0$, l'equazione $g(x,y,z) = 0$ definisce $z = f(x,y)$ in un intorno di (x_0, y_0); dalla (3.21):

$$z_x(x_0, y_0) = -\frac{g_x(x_0, y_0, z_0)}{g_z(x_0, y_0, z_0)}$$

$$z_y(x_0, y_0) = -\frac{g_y(x_0, y_0, z_0)}{g_z(x_0, y_0, z_0)}$$

e quindi l'equazione del piano tangente, scritto in forma simmetrica rispetto alle tre variabili è:

$$g_x(x_0,y_0,z_0)(x-x_0)+g_y(x_0,y_0,z_0)(y-y_0)+g_z(x_0,y_0,z_0)(z-z_0)=0. \quad (3.23)$$

***Esempio* 3.14** - Per l'*ellissoide*

$$g(x,y,z)=\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}-1=0$$

abbiamo:

$$g_x=\frac{2x}{a^2}\ ,\ g_y=\frac{2y}{b^2}\ ,\ g_z=\frac{2z}{c^2}\ .$$

L'equazione del piano tangente in un suo punto (x_0,y_0,z_0) è

$$\frac{x_0}{a^2}(x-x_0)+\frac{y_0}{b^2}(y-y_0)+\frac{z_0}{c^2}(z-z_0)=0\ ;$$

essendo $\frac{x_0^2}{a^2}+\frac{y_0^2}{b^2}+\frac{z_0^2}{c^2}=1$ questa equazione si riduce a

$$\frac{xx_0}{a^2}+\frac{yy_0}{b^2}+\frac{zz_0}{c^2}-1=0\ .$$

Passiamo ora alla generalizzazione delle nozioni di insieme di livello e di inviluppo. Ci limiteremo per semplicità al caso $n=2$. L'allievo non avrà difficoltà ad estendere le considerazioni che seguono al caso $n>2$.

Possiamo interpretare l'insieme $E_0=\{(x,y,z)\in\mathbb{R}^3\ ,\ g(x,y,z)=0\}$ come l'intersezione del grafico di $w=g(x,y,z)$ con l'iperpiano $w=0$; più in generale possiamo introdurre la famiglia di insiemi

$$E_c=\{(x,y,z)\in\mathbb{R}^3\ ,\ g(x,y,z)=c\}\qquad (c\in\mathbb{R})$$

proiezione sull'iperpiano $w=0$ dell'intersezione tra l'iperpiano $w=c$ e il grafico di $w=g(x,y,z)$.

Chiameremo E_c *insiemi di livello* di g. Se g è di classe C^1 e $\nabla g\neq\mathbf{0}$ in ogni punto E_c è ragionevole chiamare E_c *superficie di livello*.

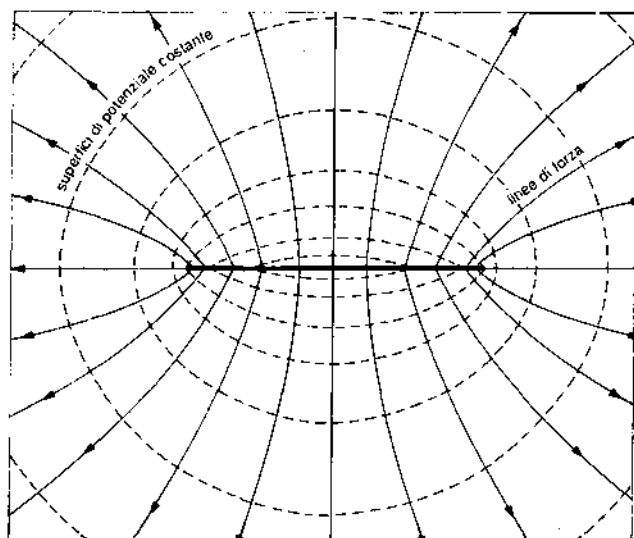



Fig. 7.32 Sezioni di superfici equipotenziali e linee di forza generate da un disco caricato uniformemente.

La formula (3.23) indica allora che il vettore ∇g è ortogonale al vettore $(x - x_0, y - y_0, z - z_0)$ (*) che giace sul piano tangente alla superficie di livello nel punto (x_0, y_0, z_0). In altri termini, *∇g è ortogonale alle superfici di livello.*

Un esempio notevole in elettrostatica è dato dalle superfici di livello $V(x, y, z) = c$ dove V rappresenta il potenziale elettrostatico. Queste superfici si chiamano *equipotenziali*. Il campo elettrostatico è $\mathbf{E} = \nabla V$ che, per quanto detto sopra, è diretto ortogonalmente alle superfici equipotenziali. In termini più fisici: *le linee di forza di* $\mathbf{E}$ *sono ortogonali alle superfici equipotenziali.*

Concludiamo il paragrafo con un cenno agli inviluppi di superfici.

Consideriamo prima il caso di una famiglia ad un parametro $\Psi(x, y, z, c) = 0$, dove (x, y, z) varia in un aperto A di $\mathbb{R}^3$ e c in un intervallo $I \subseteq \mathbb{R}$ e dove $\Psi \in C^1(A \times I)$, con $\nabla\Psi \neq \mathbf{0}$ in $A \times I$.

Definizione 3.3 - *Una superficie E è inviluppo della famiglia $\Psi(x, y, z, c) = 0$ se è tangente ad ogni superficie della famiglia lungo una curva e se queste curve di contatto costituiscono una famiglia ad un parametro, la cui unione è uguale ad E.*

Ad esempio, la famiglia di sfere unitarie con centro sull'asse z ha come inviluppo il cilindro $x^2 + y^2 = 1$ (cfr. esempio 3.15).

L'analogo della proposizione 3.4 è la seguente, di cui omettiamo la dimostrazione:

(*) Se $\Delta\mathbf{r} = (x - x_0, y - y_0, z - z_0)$, la (3.23) equivale a $< \nabla g(x_0, y_0, z_0), \Delta\mathbf{r} >= 0$.

Proposizione 3.6 - *Se E è inviluppo della famiglia $\Psi(x,y,z,c)=0$ allora la sua equazione si ottiene eliminando c tra le equazioni*

$$\Psi(x,y,z,c)=0 \qquad \Psi_c(x,y,z,c)=0\ . \tag{3.24}$$

Dal punto di vista pratico, per trovare l'eventuale inviluppo di una famiglia di superfici, si risolve il sistema (3.24), verificando poi di aver trovato effettivamente l'inviluppo.

***Esempio* 3.15** - Sia $\Psi(x,y,z,c)=x^2+y^2+(z-c)^2-1=0$ la famiglia di sfere unitarie con centro sull'asse z.

Risolvendo il sistema

$$\begin{cases} x^2+y^2+(z-c)^2-1=0 \\ -2(z-c)=0 \end{cases}$$

si trova $x^2+y^2=1$ che è l'equazione del cilindro circolare retto avente per asse l'asse z.

Le curve di contatto sono i *cerchi* unitari, paralleli al piano xy, giacenti sul piano $z=c$, centrati in $(0,0,c)$.

Nel caso delle superfici è anche possibile definire la nozione di inviluppo per una famiglia dipendente da 2 parametri c_1 e c_2:

$$\Psi(x,y,z,c_1,c_2)=0\ ,$$

come ad esempio, la famiglia

$$(x-c_1)^2+(y-c_2)^2+z^2-1=0 \tag{3.25}$$

di sfere con centro sul piano $z=0$ e raggio unitario.

Definizione 3.4 - *Una superficie E è inviluppo di una famiglia a due parametri se, in ogni suo punto, E è tangente ad una ed una sola superficie della famiglia.* (Ogni superficie della famiglia è tangente ad E in un punto e non lungo una curva come nella definizione 3.2).

L'analogo delle proposizioni 3.4 e 3.6 è la seguente di cui, ancora, omettiamo la dimostrazione.

Proposizione 3.7 - *Se E è l'inviluppo della famiglia $\Psi(x,y,z,c_1,c_2)=0$ allora la sua equazione si trova eliminando c_1 e c_2 dalle equazioni*

$$\begin{aligned} &\Psi(x,y,z,c_1,c_2)=0\ ,\\ &\Psi_{c_1}(x,y,z,c_1,c_2)=0\ ,\\ &\Psi_{c_2}(x,y,z,c_1,c_2)=0\ . \end{aligned} \tag{3.26}$$

Esempio **3.16** - Verifichiamo se la famiglia (3.25) ha inviluppo.
Risolvendo il sistema:

$$\begin{cases} (x-c_1)^2+(y-c_2)^2+z^2-1=0 \\ -2(x-c_1)=0 \\ -2(y-c_2)=0 \end{cases}$$

si trova $z^2-1=0$ ovvero i due piani $z=\pm 1$ che infatti risultano inviluppo della famiglia.

La nozione di inviluppo di superfici interviene nel celebre *principio di Huygens* in ottica geometrica:

se $\sum$ è in un certo istante una superficie d'onda, ogni suo punto si può considerare come una sorgente di onde sferiche secondarie; se v è la velocità dell'onda, queste, dopo un tempo t avranno raggio vt e il loro inviluppo costituisce la nuova superficie d'onda $\sum_t$.

Fig. 7.33

In economia, un interessante impiego della nozione di inviluppo si ha nella *teoria del costo medio di lungo periodo.*

Supponiamo che un'azienda disponga di tre diversi metodi di fabbricazione di un unico prodotto x (x è tradotto in quantità di denaro). Ogni modo di produzione avrà una curva di costo: $c_1(x), c_2(x), c_3(x)$.

La curva non tratteggiata in fig. 7.34 rappresenta la curva di minimo costo che simula la strategia di produzione dell'azienda: fino alla quantità x_1 conviene produrre col metodo 1, da x_1 ad x_2 col metodo 2, da x_2 in poi col metodo 3.

La funzione costo sarà dunque $c(x)=\min(c_1(x), c_2(x), c_3(x))$.

Se ora supponiamo che l'azienda abbia a sua disposizione un numero arbitrariamente grande di metodi di produzione per x la funzione costo sarà costituita dall'inviluppo della famiglia delle funzioni costo associate a ciascun metodo di produzione.

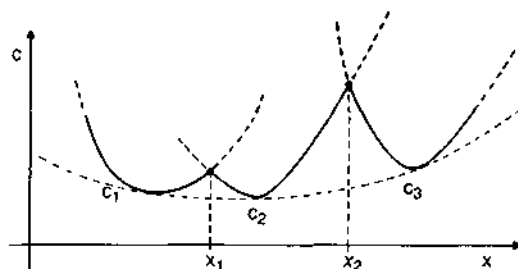



Fig. 7.34

3.6 Funzioni definite da un sistema di equazioni

Presentiamo in questo paragrafo il teorema sulle funzioni implicite nel suo aspetto generale. Il problema è di stabilire condizioni sotto le quali in un sistema del tipo seguente

$$\begin{cases} g_1(x_1,x_2,\ldots,x_n,y_1,y_2,\ldots,y_m) & = 0 \\ g_2(x_1,x_2,\ldots,x_n,y_1,y_2,\ldots,y_m) & = 0 \\ \vdots & \vdots \\ g_m(x_1,x_2,\ldots,x_n,y_1,y_2,\ldots,y_m) & = 0 \end{cases} \tag{3.27}$$

sia possibile esplicitare le m variabili $y_1,\ldots,y_m$ in funzione delle altre, $x_1,x_2,\ldots$, x_n, o, in altre parole, risolvere le m equazioni (3.27) rispetto alle m variabili $y_1,\ldots,y_m$.

Ponendo $\mathbf{x} = (x_1,x_2,\ldots,x_n)$, $\mathbf{y} = (y_1,y_2,\ldots,y_m)$, $\mathbf{g} = (g_1,g_2,\ldots,g_m)$ il sistema (3.27) si può scrivere nella forma compatta

$$\mathbf{g}(\mathbf{x},\mathbf{y}) = \mathbf{0} \tag{3.28}$$

formalmente identica alla (3.1).

Osserviamo che se $\mathbf{g}(\mathbf{x},\mathbf{y}) = \mathbf{h}(\mathbf{y})-\mathbf{x}$, il problema posto consiste nel determinare l'inversa della funzione $\mathbf{h}$ ed è già stato ampiamente trattato e risolto (almeno localmente) col teorema di inversione locale.

In quello stesso paragrafo abbiamo sottolineato che il teorema di inversione locale è l'analogo non lineare del teorema di Cramer, corrispondente al caso in cui $\mathbf{h}$ è una trasformazione lineare da $\mathbb{R}^m$ in $\mathbb{R}^m$, rappresentabile quindi con una matrice quadrata di ordine m.

Il teorema che presenteremo più avanti è invece l'analogo non lineare del teorema di Rouché–Capelli relativo ad un sistema lineare omogeneo di m equazioni in $n+m$ incognite

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n + b_{11}y_1 + \ldots + b_{1m}y_m = 0 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + b_{21}y_1 + \ldots + b_{2m}y_m = 0 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + b_{m1}y_1 + \ldots + b_{mm}y_m = 0\ . \end{cases} \tag{3.29}$$

In base al teorema di Rouché–Capelli, se la matrice A dei coefficienti di y_1, y_2, ..., y_m è non singolare allora dal sistema lineare si possono ricavare $y_1, y_2, \ldots, y_m$ in termini di $x_1, x_2, \ldots, x_n$.

Se indichiamo con $g_1, g_2, \ldots, g_m$ i primi membri del sistema (3.29), la matrice A coincide con la matrice Jacobiana di $g_1, \ldots, g_m$ rispetto a $y_1, \ldots, y_m$, che indicheremo con $\mathbf{D_y g}$:

•

$$\mathbf{D_y g} := \begin{pmatrix} \frac{\partial g_1}{\partial y_1} & \frac{\partial g_1}{\partial y_2} & \cdots & \frac{\partial g_1}{\partial y_m} \\ \frac{\partial g_2}{\partial y_1} & \frac{\partial g_2}{\partial y_2} & \cdots & \frac{\partial g_2}{\partial y_m} \\ \vdots & & & \vdots \\ \frac{\partial g_m}{\partial y_1} & \frac{\partial g_m}{\partial y_2} & \cdots & \frac{\partial g_m}{\partial y_m} \end{pmatrix} .$$

Il determinante di $\mathbf{D_y g}$ è indicato col simbolo $\frac{\partial(g_1, g_2, \ldots, g_m)}{\partial(y_1, y_2, \ldots, y_m)}$.

Il prossimo teorema afferma che se $\mathbf{D_y g}$ è non singolare in un punto, allora in un suo intorno è possibile esplicitare $\mathbf{y}$ in funzione di $\mathbf{x}$ nella (3.27).

■ **Teorema 3.8** - *Sia* $\mathbf{g} : \mathbb{R}^{n+m} \supseteq A \to \mathbb{R}^m$, *A aperto. Supponiamo che:*

i) $\mathbf{g} \in C^1(A)$

ii) nel punto $(\mathbf{x}^0, \mathbf{y}^0) \in A$ *si abbia* $\mathbf{g}(\mathbf{x}^0, \mathbf{y}^0) = \mathbf{0}$ *e*

$$\det \mathbf{D_y g}(\mathbf{x}^0, \mathbf{y}^0) \neq 0 .$$

Allora esiste un intorno V *di* $\mathbf{x}^0$ *ed un'unica funzione* $\mathbf{y} = \mathbf{f}(\mathbf{x})$, $\mathbf{f} : V \to \mathbb{R}^m$, *tale che:* $\mathbf{f} \in C^1(V)$, $\mathbf{f}(\mathbf{x}^0) = \mathbf{y}^0$ *e*

$$\mathbf{g}(\mathbf{x}, \mathbf{f}(\mathbf{x})) = \mathbf{0} \text{ per ogni } \mathbf{x} \in V . \tag{3.30}$$

Inoltre vale la formula seguente che assegna la matrice Jacobiana di $\mathbf{f}$:

$$\mathbf{Df} = -(\mathbf{D_y g})^{-1} \cdot \mathbf{D_x g} \ (^*) . \tag{3.31}$$

Dimostrazione - L'idea è di ricondursi al teorema di inversione locale risolvendo,anziché il sistema (3.27) il sistema seguente di $n + m$ equazioni in $n + m$ incognite:

$$\begin{cases} g_1(x_1, \ldots, x_n, y_1, \ldots, y_m) & = z_1 \\ g_2(x_1, \ldots, x_n, y_1, \ldots, y_m) & = z_2 \\ \vdots & \vdots \\ g_m(x_1, \ldots, x_n, y_1, \ldots, y_m) & = z_m \\ x_1 & = w_1 \\ x_2 & = w_2 \\ & \vdots \\ x_n & = w_n \end{cases} \tag{3.32}$$

(*) Più precisamente:

$$\mathbf{Df}(\mathbf{x}) = -[\mathbf{D_y g}(\mathbf{x}, \mathbf{f}(\mathbf{x}))]^{-1} \cdot \mathbf{D_x g}(\mathbf{x}, \mathbf{f}(\mathbf{x})), \quad \forall \mathbf{x} \in V .$$

Risolvere il sistema (3.32) equivale ad invertire la funzione $\mathbf{G} : (\mathbf{x},\mathbf{y}) \to (\mathbf{z},\mathbf{w})$ definita da

$$\mathbf{z} = \mathbf{g}(\mathbf{x},\mathbf{y}) \,, \quad \mathbf{w} = \mathbf{x} \,. \tag{3.33}$$

Ora, $\mathbf{G} \in C^1(A)$ e, per ipotesi, $\mathbf{G}(\mathbf{x}^0,\mathbf{y}^0) = (\mathbf{0},\mathbf{x}^0)$. Inoltre

$$\mathbf{DG} = \begin{pmatrix} \mathbf{D_x g} & \mathbf{D_y g} \\ \mathbf{I}_n & \mathbf{0}\,n,m \end{pmatrix}$$

dove $\mathbf{I}_n$ è la matrice identità di ordine n ed $\mathbf{0}\,n,m$ è la matrice nulla di ordine $n \times m$.

Poiché $\mathbf{D_y g}(\mathbf{x}^0,\mathbf{y}^0)$ è non singolare, anche $\mathbf{DG}(\mathbf{x}^0,\mathbf{y}^0)$ risulta non singolare, in quanto $\det \mathbf{DG}(\mathbf{x}^0,\mathbf{y}^0) = (-1)^{m-n} \det \mathbf{D_y g}(\mathbf{x}^0,\mathbf{y}^0)$.

Per il teorema di inversione locale esistono un intorno W di $(\mathbf{0},\mathbf{x}^0)$ ed un intorno U di $(\mathbf{x}^0,\mathbf{y}^0)$ in cui $\mathbf{G}$ è $1 \to 1$ e di classe C^1. La funzione inversa di $\mathbf{G}$ è della forma

$$\mathbf{y} = \boldsymbol{\varphi}(\mathbf{z},\mathbf{w}), \;\; \mathbf{x} = \mathbf{w} \quad \forall \mathbf{z},\mathbf{w} \in W \,. \tag{3.34}$$

Per definizione di funzione inversa, dalle (3.33), (3.34) si ottiene

$$\mathbf{z} = \mathbf{g}(\mathbf{w},\boldsymbol{\varphi}(\mathbf{z},\mathbf{w})) \quad \forall \mathbf{z},\mathbf{w} \in W \,. \tag{3.35}$$

Definiamo ora $V = \{\mathbf{x} : (\mathbf{0},\mathbf{x}) \in W\}$ e $\mathbf{f}(\mathbf{x}) = \boldsymbol{\varphi}(\mathbf{0},\mathbf{x})$.

Allora V è un intorno di $\mathbf{x}^0, \mathbf{f} \in C^1(V), \mathbf{f}(\mathbf{x}^0) = \boldsymbol{\varphi}(\mathbf{0},\mathbf{x}^0) = \mathbf{y}^0$ (da (3.34)) ed inoltre dalla (3.35):

$$\mathbf{0} = \mathbf{g}(\mathbf{x},\mathbf{f}(\mathbf{x})) \quad \forall \mathbf{x} \in V \,.$$

Per completare la dimostrazione della prima parte del teorema occorre far vedere che la funzione implicita è unica. Se infatti ce ne fossero 2, diciamo $\mathbf{f}_1$ ed $\mathbf{f}$ si avrebbe, posto $\mathbf{y}_1 = \mathbf{f}_1(\mathbf{x})$ e $\mathbf{y} = \mathbf{f}(\mathbf{x})$:

$$\begin{aligned} \mathbf{z} &= \mathbf{g}(\mathbf{x},\mathbf{f}(\mathbf{x})) = \mathbf{0} \\ \mathbf{z}_1 &= \mathbf{g}(\mathbf{x},\mathbf{f}_1(\mathbf{x})) = \mathbf{0} \end{aligned} \qquad (\mathbf{x} \in V)$$

e quindi $\mathbf{G}(\mathbf{x},\mathbf{y}) = \mathbf{G}(\mathbf{x},\mathbf{y}_1)$, che, per l'iniettività di $\mathbf{G}$ implica $\mathbf{y}_1 = \mathbf{y}$.

La formula (3.31) si ottiene dall'equazione (3.30) e dal teorema di derivazione delle funzioni composte:

$$\mathbf{D_y g}(\mathbf{x},\mathbf{f}(\mathbf{x})) \cdot \mathbf{Df}(\mathbf{x}) + \mathbf{D_y g}(\mathbf{x},\mathbf{f}(\mathbf{x})) = \mathbf{0} \quad \forall \mathbf{x} \in V \,. \;\square$$

Invitiamo il lettore a verificare che i teoremi 3.1 e 3.5 risultano casi particolari del teorema 3.8.

Anche in questo caso $\mathbf{g} \in C^k(A) \Rightarrow \mathbf{f} \in C^k(V), \mathbf{g} \in C^\infty(A) \Rightarrow \mathbf{f} \in C^\infty(V)$; differenziali successivi e formula di Taylor con resto di Peano si possono poi trovare con metodi perfettamente analoghi a quelli dell'esempio 3.13.

***Esempio* 3.17** - Sia dato il sistema

$$\begin{cases} x + \log y + 5z - 10 = 0 \\ 2x + y^2 + 3z^3 - 25 = 0 \end{cases} \tag{3.36}$$

del quale $(0,1,2)$ è una soluzione.

Le funzioni $g_1(x,y,z) = x + \log y + 5z - 10$, $g_2(x,y,z) = 2x + y^2 + 3z^3 - 25$ sono di classe C^∞ in un intorno di $(0,1,2)$.

Ci chiediamo quale coppia di variabili sia esplicitabile rispetto all'altra in tale intorno.

Abbiamo:

$$\nabla g_1 = (1, \frac{1}{y}, 5), \ \ \nabla g_2 = (2, 2y, 9z^2)$$

e, di conseguenza:

$$\frac{\partial(g_1,g_2)}{\partial(x,y)}(0,1,2) = \begin{vmatrix} 1 & 1 \\ 2 & 2 \end{vmatrix} = 0$$

$$\frac{\partial(g_1,g_2)}{\partial(x,z)}(0,1,2) = \begin{vmatrix} 1 & 5 \\ 2 & 36 \end{vmatrix} = 26$$

$$\frac{\partial(g_1,g_2)}{\partial(y,z)}(0,1,2) = \begin{vmatrix} 1 & 5 \\ 2 & 36 \end{vmatrix} = 26 \ .$$

Dal teorema 3.8 deduciamo che dal sistema (3.36) è possibile esplicitare x e z in funzione di y per y in un intorno di 1, oppure y, z in funzione di x, per x in un intorno di 0.

Scegliamo l'ultima possibilità, dunque $\mathbf{f}(x) = (y(x), z(x))$, e proponiamoci di scrivere la formula di Mac Laurin con il resto di Peano, arrestata al 2° ordine per $\mathbf{f}$. Ricordiamo che $y(0) = 1, z(0) = 2$. Essendo $\mathbf{D}_x\mathbf{g}(x,y,z) = \binom{1}{2}$, dalla formula (3.31) si ha:

$$\mathbf{Df}(0) = \begin{pmatrix} y'(0) \\ z'(0) \end{pmatrix} = -\begin{pmatrix} 1 & 5 \\ 2 & 36 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} = -\frac{1}{26}\begin{pmatrix} 36 & -5 \\ -2 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} =$$
$$= -\frac{1}{26}\begin{pmatrix} 26 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \end{pmatrix} .$$

Dunque $y'(0) = -1, \ z'(0) = 0$.

Alla stessa conclusione si poteva pervenire derivando rispetto ad x le due equazioni del sistema, pensando $y = y(x)$ e $z = z(x)$:

$$\begin{cases} 1 + \dfrac{y'}{y} + 5z' = 0 \\ 2 + 2yy' + 9z^2z' = 0 \ . \end{cases} \tag{3.37}$$

Calcolando in $x = 0$ si ottiene,

$$\begin{cases} y'(0) + 5z'(0) = -1 \\ 2y'(0) + 36z'(0) = -2 \end{cases}$$

che assegna appunto i valori precedentemente trovati.

Derivando nuovamente, sempre rispetto ad x le (3.37) otteniamo:

$$\begin{cases} \dfrac{y''y-(y')^2}{y^2}+5z''=0 \\ 2(y')^2+2yy''+18z(z')^2+9z^2z''=0 \ . \end{cases}$$

Calcolando in $x=0$:

$$\begin{cases} y''(0)+5z''(0)=1 \\ 2y''(0)+36z''(0)=-2 \end{cases}$$

da cui si ricava $y''(0)=\dfrac{23}{13}$, $z''(0)=-\dfrac{2}{13}$.

Conclusione:

$$y(x)=1-x+\frac{23}{26}x^2+o(x^2)$$

$$z(x)=2-\frac{1}{13}x^2+o(x^2) \ .$$

Esercizi

1. Verificare che l'equazione $e^{x-y}+x^2-y^2-e(x+1)+1=0$ definisce implicitamente $y=f(x)$ in un intorno di $x=0$ con $f(0)=-1$.
Dimostrare che $x=0$ è punto di minimo relativo per f.

2. Dimostrare il seguente teorema globale sulle funzioni implicite.
Sia $g:\mathbb{R}^2\to\mathbb{R}$ tale che:

i) $g\in C(\mathbb{R}^2)$, esiste $g_y(x,y), \forall(x,y)\in\mathbb{R}^2$ ed è positiva (risp. negativa).

ii) $\lim\limits_{y\to+\infty} g(x,y)>0$ (risp. <0) e $\lim\limits_{y\to-\infty} g(x,y)<0$ (risp. >0), $\forall x\in\mathbb{R}$.

Allora esiste una ed una sola funzione $f:\mathbb{R}\to\mathbb{R}$ tale che $g(x,f(x))=0, \forall x\in\mathbb{R}$. Se inoltre $g\in C^k(\mathbb{R})$ anche $f\in C^k(\mathbb{R}), k\in\mathbb{N}$ oppure $k=\infty$.

3.* Sia data l'equazione $x^2y+e^{x+y}=0$.
Dimostrare che essa definisce una funzione $y=f(x)$ per $x\in\mathbb{R}\setminus\{0\}$.
Verificare inoltre che:

a) $f(x)<0 \ \forall x\in\mathbb{R}\setminus\{0\}$.

b) $f(x)\to 0_-$ se $x\to-\infty$ e quindi f ha $y=0$ come asintoto orizzontale per $x\to-\infty$.

c) $f(x)\to-\infty$ per $x\to 0$ e quindi f ha $x=0$ come asintoto verticale.

d) $f(x)\to-\infty$ per $x\to+\infty$, ma non esiste asintoto obliquo.

e) $f(x)$ ha un massimo locale per $x=2$.

Tracciare un grafico qualitativo di $y=f(x)$.

4. Studiare gli insiemi di livello delle seguenti funzioni:

a) $g(x,y) = \varphi(x - y^2)$;

b) $g(x,y) = e^{xy}$;

c) $g(x,y) = \log(|\operatorname{tg} y - x|)$.

5. Verificare che l'equazione $\operatorname{arctg} z + xy^2 + xz - y^3 - 1 = 0$ definisce $z = z(x,y)$ in un intorno del punto $(0,-1,0)$. Calcolare l'equazione del piano tangente al grafico di $z = z(x,y)$ in questo punto.

6. Sia $g = g(x,y,z)$ differenziabile e positivamente omogenea di grado $\alpha \geq 0$, in $\mathbb{R}^3$. Dimostrare che il piano tangente alla superficie di livello $g(x,y,z) = 1$ in un punto (x_0, y_0, z_0) ha equazione

$$xg_x(x_0,y_0,z_0) + yg_y(x_0,y_0,z_0) + zg_z(x_0,y_0,z_0) = \alpha\ .$$

7. Le due superfici di livello $F(x,y,z) = c_1$ e $G(x,y,z) = c_2$ si intersecano in un punto (x_0,y_0,z_0), non singolare per entrambe.

Detto ω l'angolo tra le normali al piano tangente alle due superfici in (x_0,y_0,z_0), verificare che

$$\cos\omega = \frac{< \nabla F(x_0,y_0,z_0), \nabla G(x_0,y_0,z_0) >}{\|\nabla F(x_0,y_0,z_0)\| \cdot \|\nabla G(x_0,y_0,z_0)\|}\ .$$

8. Sia f omogenea di grado zero in $\mathbb{R}^2$ e g radiale (cioè $g(x,y) = g(\sqrt{x^2+y^2})$). Descrivere gli insiemi di livello di f e g.

9. Dimostrare che se $x = x(c)$ e $y = y(c)$ è soluzione di (3.16) e $\Psi_{cc} \neq 0$ allora $x'(c)$ e $y'(c)$ non possono essere entrambe nulle.
[Suggerimento: derivare rispetto a c, $\Psi_c(x(c), y(c), c) = 0 \ldots$].

10. Dimostrare che l'astroide è inviluppo anche della famiglia di rette

$$\frac{x}{\cos c} + \frac{y}{\sin c} = 1\ .$$

11. Studiare la natura del punto singolare $(0,0)$ per la curva algebrica seguente (cardioide)

$$(x^2 + y^2 - ax)^2 - l^2(x^2 + y^2) = 0 \quad (l \neq 0)\ .$$

12. Dimostrare che l'equazione polare di una conica, rispetto ad un polo e ad un asse polare scelti opportunamente, è della forma

$$\rho = \frac{p}{1 + e\cos\theta}\ .$$

Interpretare p ed e.

13. Scrivere le equazioni polari della lemniscata, del folium di Cartesio, della cardioide, della cissoide e delle curve negli esempi 3.7, 3.8.

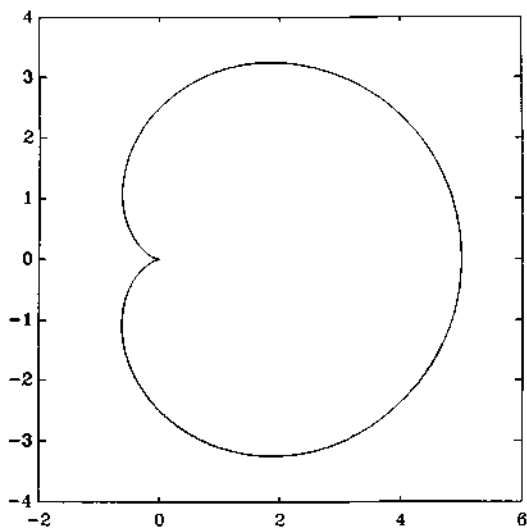



Fig. 7.35 Cardioide ($a = l = 2.5$).

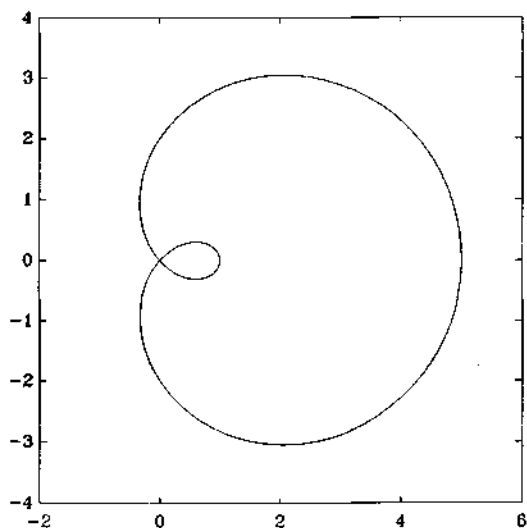



Fig. 7.36 Cardioide ($a = 3,\ l = 2$).

14. Studiare le singolarità delle seguenti curve algebriche:

$$a(x-a)(x^2+y^2) - k^2x^2 = 0 \quad a, k \neq 0 \quad \text{(Concoide di Sluse)}$$

$$(y^4 - 4m^2)(x^2+y^2) + 4m^2 = 0 \quad m \neq 0 \quad \text{(Trisecante di Delanges)}\ .$$

15. Sia $\mathbf{F} : \mathbb{R}^4 \to \mathbb{R}^2$ definita da:

$$\mathbf{F}(x_1, x_2, y_1, y_2) = \\ = ((x_2+3)y_1 - \operatorname{tg}(y_1+y_2) + 2x_1 \ , \ \sin(y_1+y_2) + 3x_2 - x_1(y_2+3)) \ .$$

Dimostrare che $\mathbf{F} = 0$ definisce implicitamente $y_1 = y_1(x_1, x_2)$ e $y_2 = y_2(x_1, x_2)$ in un intorno di $(0,0,0,0)$.

Calcolare $\dfrac{\partial y_1}{\partial x_2}(0,0)$.

16. L'equazione $g(x, y) = 0$ definisce $y = f(x), x \in (0, +\infty)$.

È vero che se $f \to +\infty$ per $x \to +\infty$ e se $g(x, mx+q) \to 0$ la retta $y = mx + q$ è asintoto obliquo per f?

17.** Completare i dettagli della dimostrazione della proposizione 3.3.

[Suggerimento: a) posto $\Delta y_1 = y_1(x) - y_1(x_0)$, $\Delta x = x - x_0$, usare la formula di Taylor per ricavare $\dfrac{\Delta y_1}{\Delta x} = -[g_{yx}(\overline{x}, \overline{y}) \pm \sqrt{-H(\overline{x}, \overline{y})}]/g_{yy}(\overline{x}, \overline{y})$ dove $(\overline{x}, \overline{y})$ è sul segmento $(x, y(x)), (x_0, y_0)$. b) Usare la (3.14) per dimostrare che in a) deve valere il segno $-$, definitivamente per $x \to x_0$. c) Dedurre che in x_0, y_1 è derivabile e dare la formula per la derivata $y_1'(x_0)$. Ragionare in maniera analoga per y_2. d) Se $x \neq x_0$, essendo $g_y(x, y_1(x)) \neq 0$, si può ragionare come nel teorema 3.1. e) Dalla formula per $y_1'(x)$ dedurre che y_1' è continua].

Integrali di funzioni di una variabile
Serie numeriche 8

Dopo aver introdotto l'operazione di *limite* (Cap. 4) e successivamente quella di *derivazione* (Cap. 6), che ne è uno sviluppo diretto, considereremo, in questo capitolo, la terza fondamentale operazione del calcolo infinitesimale: l'*integrazione*. Va osservato che, dal punto di vista storico, i tre concetti si sono susseguiti esattamente nel verso opposto: gli studi sulla "quadratura delle figure piane" (legati al concetto di integrale) si possono far risalire all'antichità, ad esempio ad Archimede ed al suo metodo detto di "esaustione"; il problema di determinare la tangente ad una curva (e la conseguente invenzione delle derivate ad opera di Leibniz e Newton) comincia solo nella prima metà del seicento (con Fermat e Descartes); l'operazione di limite poi, come abbiamo già avuto modo di accennare, introdotta da D'Alembert (ma siamo già nel '700), venne poi sviluppata potentemente da Cauchy e da Weierstrass.

Tra le varie versioni dell'integrale di una funzione abbiamo scelto quella dovuta a Bernard Riemann ed è presentata nella sezione 1.

Nella sezione 2 viene presentata la teoria delle serie numeriche, che precisa il concetto di somma di un numero infinito di termini. Sebbene le serie, come vedremo, non sono nulla di diverso, sostanzialmente, dalle successioni, la presenza della somma conferisce loro delle forti analogie con gli integrali, soprattutto con quelli generalizzati, oggetto della sezione 3 insieme con altre estensioni del concetto di integrale (integrale di Stieltjes) che trovano applicazione ad esempio in statistica e calcolo delle probabilità.

1. INTEGRALE DI RIEMANN

1.1 Definizione di integrale

Definizione 1.1 - *Sia* $[a,b], a < b$, *un intervallo limitato. Un insieme finito* $\{x_i\}(i = 0, 1, \ldots, n)$ *di punti di questo intervallo tali che:* $x_0 = a < x_1 < \ldots <$

$< x_{n-1} < x_n = b$ *viene detto suddivisione di* $[a,b]$.

Indicheremo una tale suddivisione con $\mathcal{D}(x_0,\ldots,x_n)$, o semplicemente $\mathcal{D}$. L'insieme delle suddivisioni può essere parzialmente ordinato con l'*ordinamento* introdotto in $\mathcal{P}([a,b])$ (la famiglia dei sottoinsiemi di $[a,b]$) dalla relazione $\subseteq$; diremo cioè che una suddivisione $\mathcal{D}_1$ è *meno fine* della suddivisione $\mathcal{D}_2$ se $\mathcal{D}_1 \subseteq \mathcal{D}_2$, cioè se $\mathcal{D}_2$ possiede, oltre ai punti di $\mathcal{D}_1$, anche un solo punto in più. Inoltre, date due suddivisioni qualsiansi $\mathcal{D}_1$ e $\mathcal{D}_2$ possiamo sempre considerarne una terza, $\mathcal{D}_3$, più fine di entrambe: per esempio, $\mathcal{D}_3 = \mathcal{D}_1 \cup \mathcal{D}_2$.

Consideriamo ora una funzione $f : [a,b] \to \mathbb{R}$ limitata:

$$m \leq f(x) \leq M \qquad \forall x \in [a,b] \,. \tag{1.1}$$

Sia $\mathcal{D} = \mathcal{D}(x_0,\ldots,x_n)$ una suddivisione di $[a,b]$. Poniamo, per $i = 1,2,\ldots,n$,

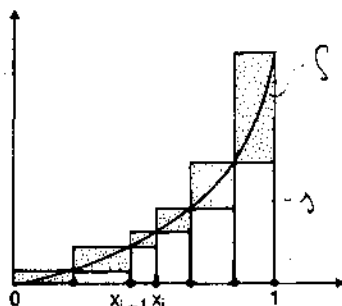
$$m_i = \inf_{(x_{i-1},x_i)} f \qquad M_i = \sup_{(x_{i-1},x_i)} f \,.$$

Poniamo anche, scrivendo $\Delta x_i = x_i - x_{i-1}$,

$$s = s(\mathcal{D}, f) = \sum_{i=1}^{n} m_i \Delta x_i \tag{1.2}$$

$$\mathcal{S} = \mathcal{S}(\mathcal{D}, f) = \sum_{i=1}^{n} M_i \Delta x_i \,. \tag{1.3}$$

Chiameremo $s(\mathcal{D}, f)$ e $\mathcal{S}(\mathcal{D}, f)$ rispettivamente *somma inferiore* e *somma superiore* di f relativa alla suddivisione $\mathcal{D}$.

La figura 8.1 illustra il significato geometrico di s e $\mathcal{S}$ nel caso della parabola: $y = x^2$ per l'intervallo $[0,1]$.

$$s = \sum_{i=1}^{n} x_{i-1}^2 (x_i - x_{i-1})$$

$$\mathcal{S} = \sum_{i=1}^{n} x_i^2 (x_i - x_{i-1}) \,.$$

Fig. 8.1

Osserviamo subito che, per *ogni* suddivisione $\mathcal{D}$ di $[a,b]$, risulta, per la (1.1)

$$m(b-a) \leq s(\mathcal{D}, f) \leq S(\mathcal{D}, f) \leq M(b-a) \ . \tag{1.4}$$

Sono perciò *finite* le due quantità seguenti:

$$\inf_{\mathcal{D}} S(\mathcal{D}, f) \qquad \sup_{\mathcal{D}} s(\mathcal{D}, f)$$

cioè l'estremo inferiore (preso sull'insieme delle suddivisioni di $[a,b]$) delle somme superiori e l'estremo superiore delle somme inferiori.

Notiamo ora che la (1.4), che confronta la somma inferiore e la somma superiore di f relative alla *stessa* suddivisione, rimane vera anche se le suddivisioni sono diverse; infatti vale il seguente

Lemma 1.1 - *Se $\mathcal{D}_1$ e $\mathcal{D}_2$ sono due suddivisioni di $[a,b]$ e $\mathcal{D}_1 \subseteq \mathcal{D}_2$ risulta, per una funzione limitata $f : [a,b] \to \mathbb{R}$:*

$$s(\mathcal{D}_1, f) \leq s(\mathcal{D}_2, f) \tag{1.5}$$

$$S(\mathcal{D}_1, f) \geq S(\mathcal{D}_2, f) \ . \tag{1.6}$$

Dimostrazione - Supponiamo dapprima che $\mathcal{D}_2$ contenga esattamente un punto: $\overline{x}$,in più di $\mathcal{D}_1 = \mathcal{D}_1(x_0 \dots x_n)$; sarà $x_{i-1} < \overline{x} < x_i$ per un certo i $(1 \leq i \leq n)$. Posto

$$\mu_1 = \inf_{(x_{i-1}, \overline{x})} f \qquad \mu_2 = \inf_{(\overline{x}, x_i)} f$$

risulta: $\mu_1 \geq m_i$, $\mu_2 \geq m_i$ e inoltre

$$\begin{aligned} s(\mathcal{D}_2, f) - s(\mathcal{D}_1, f) &= \mu_1(\overline{x} - x_{i-1}) + \mu_2(x_i - \overline{x}) - m_i(x_i - x_{i-1}) = \\ &= (\mu_1 - m_i)(\overline{x} - x_{i-1}) + (\mu_2 - m_i)(x_i - \overline{x}) \geq 0 \ . \end{aligned} \tag{1.7}$$

Ripetendo questo stesso ragionamento più volte si dimostra la (1.5) nel caso generico che $\mathcal{D}_2$ contenga più punti rispetto a $\mathcal{D}_1$. Analogamente si dimostra la (1.6). □

Osservazione 1.1 - Dalla (1.7), posto $\delta_1 = \max_{i=1,\dots,n} \Delta x_i$ e $\overline{M} = \sup_{[a,b]} |f|$, e osservando che $\mu_1 - m_i \leq 2\overline{M}$, $\mu_2 - m_i \leq 2\overline{M}$, si ricava anche: $s(\mathcal{D}_2, f) - s(\mathcal{D}_1, f) \leq 4\overline{M}\delta_1$. Ripetendo il ragionamento nel caso che $\mathcal{D}_2$ contenga esattamente k punti in più di $\mathcal{D}_1$ si trova:

$$s(\mathcal{D}_2, f) - s(\mathcal{D}_1, f) \leq 4k\overline{M}\delta_1 \tag{1.8}$$

e, analogamente,

$$S(\mathcal{D}_1, f) - S(\mathcal{D}_2, f) \leq 4k\overline{M}\delta_1 \ . \tag{1.9}$$

Dal Lemma 1.1 ricaviamo il seguente

Corollario 1.2 - *Per ogni funzione* $f : [a,b] \to \mathbb{R}$ *limitata risulta*:

$$\inf_{\mathcal{D}} \mathcal{S}(\mathcal{D}, f) \geq \sup_{\mathcal{D}} s(\mathcal{D}, f) . \tag{1.10}$$

Dimostrazione - Se $\mathcal{D}_1$ e $\mathcal{D}_2$ sono due qualsiansi suddivisioni di $[a, b]$, sia $\mathcal{D}_3 = \mathcal{D}_1 \cup \mathcal{D}_2$; risulta, come abbiamo già osservato, $\mathcal{D}_3$ più fine di $\mathcal{D}_1$ e di $\mathcal{D}_2$. Perciò, applicando il Lemma 1.1, abbiamo

$$s(\mathcal{D}_1, f) \leq s(\mathcal{D}_3, f) \leq \mathcal{S}(\mathcal{D}_3, f) \leq \mathcal{S}(\mathcal{D}_2, f) ; \tag{1.11}$$

abbiamo perciò un confronto tra somme inferiori e superiori di f relative a suddivisioni diverse. Dalla (1.11) segue subito la (1.10). □

Ora, come sarà chiaro dai prossimi esempi, per una assegnata funzione f due circostanze possono verificarsi:

o $\quad \sup_{\mathcal{D}} s = \inf_{\mathcal{D}} \mathcal{S}$

o $\quad \sup_{\mathcal{D}} s < \inf_{\mathcal{D}} \mathcal{S}$.

Definizione 1.2 - *Una funzione* $f : [a, b] \to \mathbb{R}$ *limitata si dice integrabile secondo Riemann se succede che:* $\sup_{\mathcal{D}} s = \inf_{\mathcal{D}} \mathcal{S}$; *il valore comune di questi due estremi si chiama* **integrale** *(di Riemann) di f in* $[a, b]$ *e si denota con uno dei simboli seguenti*:

$$\int_a^b f(x)dx \quad , \quad \int_I f(x)dx \quad , \quad \int_a^b f \quad , \quad \mathcal{I}(I, f)$$

dove $I = [a, b]$ è il *dominio di integrazione*, e $f = f(x)$ è la *funzione integranda*.

Scriveremo anche semplicemente $\mathcal{I}(f)$ quando sia chiaro dal contesto quale sia l'intervallo I. Osserviamo esplicitamente che la variabile di integrazione, indicata con x nella scrittura precedente, è "muta", come l'indice di sommatoria in una somma, e perciò può essere sostituita con una lettera qualsiasi, anche nel corso di uno stesso calcolo, dal momento che non compare nel risultato finale.

L'insieme delle funzioni limitate integrabili secondo Riemann su un intervallo $I = [a, b]$ si denoterà con $\mathcal{R}(I)$ o $\mathcal{R}(a, b)$.

Mostriamo subito che tale insieme non è vuoto, ma nemmeno coincide con l'insieme delle funzioni limitate in $[a, b]$.

***Esempio* 1.1** - Ogni costante c è integrabile su qualunque intervallo $[a, b]$ e risulta:

$$\int_a^b c\,dx = c(b - a) .$$

Infatti, per qualunque suddivisione $\mathcal{D}$ di $[a,b]$, abbiamo:

$$s = S = \sum_{i=1}^{n} c\Delta x_i = c\sum_{i=1}^{n}(x_i - x_{i-1}) = c(b-a)\ .$$

***Esempio* 1.2** - Sia $f : [a,b] \to \mathbb{R}$ la funzione di Dirichlet ($f(x) = 1$ se $x \in \mathbb{Q} \cap [a,b]$, $f(x) = 0$ se $x \in [a,b] \setminus \mathbb{Q}$). Allora, per qualunque suddivisione $\mathcal{D}$ di $[a,b]$ risulta:

$$s = \sum_{i=1}^{n} 0 \cdot \Delta x_i = 0$$

$$S = \sum_{i=1}^{n} 1 \cdot \Delta x_i = (b-a)\ .$$

Perciò f non è integrabile secondo Riemann.

D'ora in avanti parleremo semplicemente di funzioni *integrabili* o no, senza la precisazione "secondo Riemann".

A questo punto si pongono naturalmente due questioni:

1) Quali classi di funzioni limitate sono integrabili?

2) Se $f \in \mathcal{R}(a,b)$, il calcolo di $\int_a^b f$ mediante la definizione risulta troppo complicato. Vi sono altri metodi di calcolo?

Alla questione 1) si dà una risposta parziale nel paragrafo 1.3. Nel paragrafo 1.5 è invece presentato il teorema fondamentale che, collegando il calcolo differenziale con quello integrale, permette il calcolo di un integrale col metodo detto *"per variazione di una primitiva"*.

È però necessario introdurre prima due caratterizzazioni dell'integrale e metterne in evidenza alcune proprietà.

1.2 Caratterizzazioni dell'integrale e significato geometrico

■ **Teorema 1.3** - *Una funzione $f : [a,b] \to \mathbb{R}$ limitata appartiene a $\mathcal{R}(a,b)$ se e solo se, $\forall \varepsilon > 0$ si può trovare una suddivisione $\mathcal{D}_\varepsilon$ (dipendente da ε) di $[a,b]$ tale che:*

$$S(\mathcal{D}_\varepsilon, f) - s(\mathcal{D}_\varepsilon, f) < \varepsilon\ . \tag{1.12}$$

Dimostrazione - *i*) Valga la (1.12). Risulta allora

$$0 \le \inf_{\mathcal{D}} S - \sup_{\mathcal{D}} s \le S(\mathcal{D}_\varepsilon, f) - s(\mathcal{D}_\varepsilon, f) < \varepsilon\ .$$

Per l'arbitrarietà di ϵ segue che $\inf_{\mathcal{D}} S = \sup_{\mathcal{D}} s$, cioè $f \in \mathcal{R}(a,b)$.

ii) Viceversa, sia $f \in \mathcal{R}(a,b)$, cioè

$$\inf_{\mathcal{D}} \mathcal{S} = \sup_{\mathcal{D}} s = \mathcal{I}(f) .$$

Fissato ε, esiste una suddivisione $\mathcal{D}_1$ (dipendente da ε) tale che:

$$\mathcal{I}(f) - s(\mathcal{D}_1, f) < \frac{\varepsilon}{2}$$

e una suddivisione $\mathcal{D}_2$ (pure dipendente da ε) tale che:

$$\mathcal{S}(\mathcal{D}_2, f) - \mathcal{I}(f) < \frac{\varepsilon}{2}$$

Se $\mathcal{D}_3$ è una suddivisione più fine di $\mathcal{D}_1$ e $\mathcal{D}_2$, otteniamo

$$\mathcal{S}(\mathcal{D}_3, f) - s(\mathcal{D}_3, f) \leq \mathcal{S}(\mathcal{D}_2, f) - s(\mathcal{D}_1, f) < \varepsilon$$

e perciò la (1.12) è vera con $\mathcal{D}_\varepsilon = \mathcal{D}_3$. □

La definizione 1.2 ha introdotto l'integrale come estremo superiore/inferiore di una certa classe di numeri. Il teorema 1.3 ci porta verso una definizione dell'integrale come limite. Si intuisce infatti che l'integrale di f su $[a,b]$ dovrebbe essere il "limite" comune a cui "tendono" le somme superiori e inferiori quando la suddivisione di $[a,b]$ diventa sempre più fine. Precisiamo e generalizziamo questo punto di vista.

Definizione 1.3 - *Se $\mathcal{D}$ è una suddivisione di $[a,b]$, poniamo*

$$|\mathcal{D}| := \max_{1 \leq i \leq n} \{(x_i - x_{i-1})\} \tag{1.13}$$

e chiamiamo questo numero **ampiezza** *della suddivisione.*

Introduciamo poi, accanto alle somme inferiori e superiori, le *somme integrali* (alla Riemann):

$$\sigma(\mathcal{D}, f) = \sum_{i=1}^{n} \mu_i \Delta x_i \tag{1.14}$$

dove μ_i è un qualsiasi valore compreso tra m_i e M_i (*):

$$m_i \leq \mu_i \leq M_i .$$

Evidentemente si ha

$$s(\mathcal{D}, f) \leq \sigma(\mathcal{D}, f) \leq \mathcal{S}(\mathcal{D}, f) . \tag{1.15}$$

(*) A rigore, una notazione completa dovrebbe essere: $\sigma(\mathcal{D}, f, \{\mu_i\})$, poiché le somme σ dipendono, oltre che da $\mathcal{D}$ e da f, anche dalla scelta dei μ_i; ma è inutile appesantire la notazione.

Definizione 1.4 - *Diciamo che il numero reale l è il limite di $\sigma(\mathcal{D}, f)$ per $|\mathcal{D}| \to 0$, e scriveremo,*

$$\lim_{|\mathcal{D}|\to 0} \sigma(\mathcal{D}, f) = l$$

se, $\forall \varepsilon > 0$, si può trovare un $\delta > 0$ tale che si abbia

$$|\sigma(\mathcal{D}, f) - l| < \varepsilon$$

per ogni suddivisione $\mathcal{D}$ con $|\mathcal{D}| < \delta$ e per ogni scelta dei valori μ_i.

■ **Teorema 1.4** - *Una funzione $f : [a, b] \to \mathbb{R}$ limitata appartiene a $\mathcal{R}(a, b)$ se e solo se esiste finito $\lim_{|\mathcal{D}|\to 0} \sigma(\mathcal{D}, f)$. In tale caso risulta:*

$$\lim_{|\mathcal{D}|\to 0} \sigma(\mathcal{D}, f) = \mathcal{I}(f) \ .$$

Dimostrazione - *i*) Sia $\lim_{|\mathcal{D}|\to 0} \sigma(\mathcal{D}, f) = l \in \mathbb{R}$. Ciò significa che, fissato $\varepsilon > 0$, si può trovare un $\delta > 0$ tale che

$$l - \frac{\varepsilon}{2} < \sigma(\mathcal{D}, f) < l + \frac{\varepsilon}{2}$$

per ogni suddivisione $\mathcal{D}$ con $|\mathcal{D}| < \delta$ e per ogni scelta dei valori μ_i.

Scegliendo allora prima $\mu_i = m_i$ e poi $\mu_i = M_i$ otteniamo, dalla (1.15),

$$l - \frac{\varepsilon}{2} < s(\mathcal{D}, f) \leq S(\mathcal{D}, f) < l + \frac{\varepsilon}{2} \tag{1.16}$$

da cui

$$S(\mathcal{D}, f) - s(\mathcal{D}, f) < \varepsilon \ .$$

L'integrabilità di f segue allora dal teorema 1.3. Inoltre, dalla (1.16) ricaviamo

$$l - \frac{\varepsilon}{2} < \mathcal{I}(f) < l + \frac{\varepsilon}{2}$$

da cui, per l'arbitrarietà di ε, risulta

$$\mathcal{I}(f) = l \ .$$

ii) Sia ora $f \in \mathcal{R}(a, b)$; fissato $\varepsilon > 0$ esiste una suddivisione $\mathcal{D}_\varepsilon$ tale che

$$S(\mathcal{D}_\varepsilon, f) < \mathcal{I}(f) + \frac{\varepsilon}{2}$$

$$s(\mathcal{D}_\varepsilon, f) > \mathcal{I}(f) - \frac{\varepsilon}{2} \ .$$

Le precedenti disuguaglianze varranno anche per una qualunque suddivisione più fine di $\mathcal{D}_\varepsilon$.

Sia ora $\mathcal{D}$ una qualunque suddivisione di ampiezza $|\mathcal{D}| \leq \delta$.

Risulta, posto $\mathcal{D}_1 = \mathcal{D}_\varepsilon \cup \mathcal{D}$ ($\mathcal{D}_1$ è più fine di $\mathcal{D}_\varepsilon$ e di $\mathcal{D}$) e tenendo presente la (1.9):

$$\begin{aligned} \mathcal{S}(\mathcal{D}, f) &= \mathcal{S}(\mathcal{D}, f) - \mathcal{S}(\mathcal{D}_1, f) + \mathcal{S}(\mathcal{D}_1, f) \le \\ &\le \mathcal{S}(\mathcal{D}, f) - \mathcal{S}(\mathcal{D}_1, f) + \mathcal{S}(\mathcal{D}_\varepsilon, f) \le \\ &\le 4k\overline{M}\delta + \mathcal{I}(f) + \frac{\varepsilon}{2} \end{aligned}$$

se $\mathcal{D}_1$ contiene k punti in più di $\mathcal{D}$ e $\overline{M} = \sup\limits_{[a,b]} |f|$. Perciò se scegliamo $\delta = \varepsilon/8\overline{M}k$ risulta

$$\mathcal{S}(\mathcal{D}, f) < \mathcal{I}(f) + \varepsilon\ .$$

Analogamente si mostra che

$$s(\mathcal{D}, f) > \mathcal{I}(f) - \varepsilon$$

per ogni suddivisione con $|\mathcal{D}| < \delta$.

Allora si ha, per una qualunque somma $\sigma(\mathcal{D}, f)$ relativa ad una suddivisione $\mathcal{D}$ con $|\mathcal{D}| < \delta$:

$$\mathcal{I}(f) - \varepsilon < s(\mathcal{D}, f) \le \sigma(\mathcal{D}, f) \le \mathcal{S}(\mathcal{D}, f) < \mathcal{I}(f) + \varepsilon$$

da cui segue

$$\lim_{|\mathcal{D}| \to 0} \sigma(\mathcal{D}, f) = \mathcal{I}(f)\ . \ \square$$

Vediamo ora il significato geometrico dell'integrale.

Sia $f \in \mathcal{R}(a, b)$ non negativa. L'insieme

$$T = \{(x, y) \in \mathbb{R}^2\ ;\ a \le x \le b\ ,\ 0 \le y \le f(x)\}$$

si chiama *trapezioide* (relativo alla funzione f). In figura 8.2a è indicato il trapezioide, e il *plurirettangolo* (cioè la figura ottenuta dall'unione di un numero finito di rettangoli non sovrapponentisi) relativo ad una somma inferiore; in figura 8.2b il plurirettangolo relativo ad una somma superiore; in figura 8.2c è indicato il plurirettangolo relativo ad una somma di Riemann che non è né inferiore né superiore.

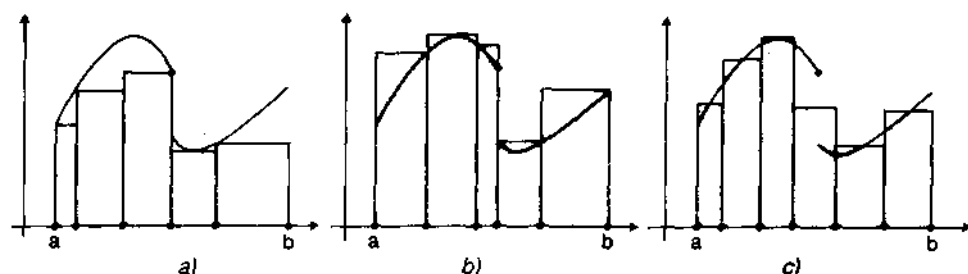



Fig. 8.2

Tutte queste somme rappresentano geometricamente l'area dei plurirettangoli corrispondenti. Il teorema 1.4 afferma che, al tendere a zero della massima ampiezza degli intervallini della suddivisione di $[a,b]$ tutte queste aree tendono ad un limite comune, che è naturale definire *area del trapezioide*. Così, attraverso l'integrale, possiamo ampliare notevolmente la classe delle figure piane per le quali è definito il concetto di misura o area.

1.3 Classi di funzioni integrabili

Esaminiamo ora alcune classi di funzioni integrabili.

▪ **Teorema 1.5** - *Ogni funzione* $f : [a,b] \to \mathbb{R}$ *continua è integrabile.*
In simboli: $C^0([a,b]) \subset \mathcal{R}(a,b)$.

Dimostrazione - Conviene richiamare che una funzione continua su un intervallo chiuso e limitato è anzitutto *limitata* ed è anche *uniformemente continua*; inoltre ammette massimo e minimo e assume ogni valore compreso tra il massimo e il minimo.

Fissato $\varepsilon > 0$, per l'uniforme continuità di f, esiste un $\delta > 0$ tale che, $\forall x', x'' \in [a,b]$ si ha:

$$|f(x') - f(x'')| < \frac{\varepsilon}{b-a} \qquad \text{se } |x' - x''| < \delta \, .$$

Sia $\mathcal{D} = \{x_0, \ldots, x_n\}$ una suddivisione di $[a,b]$ con $|\mathcal{D}| < \delta$. Abbiamo, indicando con x_i' e x_i'' rispettivamente un punto dell'intervallino $[x_{i-1}, x_i]$ in cui f assume il suo valore massimo M_i e il minimo m_i,

$$\begin{aligned} S(\mathcal{D}, f) - s(\mathcal{D}, f) &= \sum_{i=1}^{n} M_i \Delta x_i - \sum_{i=1}^{n} m_i \Delta x_i = \\ &= \sum_{i=1}^{n} [f(x_i') - f(x_i'')] \Delta x_i \le \frac{\varepsilon}{b-a} \sum_{i=1}^{n} \Delta x_i = \varepsilon \end{aligned}$$

Dal teorema 1.3 segue l'integrabilità di f. □

Osservazione 1.2 - Per una funzione continua su $[a,b]$ le somme di Riemann possono scriversi anche:

$$\sigma(\mathcal{D}, f) = \sum_{i=1}^{n} \mu_i \Delta x_i = \sum_{i=1}^{n} f(t_i) \Delta x_i$$

avendo indicato con t_i un punto dell'intervallino $[x_{i-1}, x_i]$ in cui risulti $f(t_i) = \mu_i$. La somma σ prende, in tal caso, il nome di *somma di Cauchy*. Dal teorema 1.4 scende allora che l'integrale $\mathcal{I}(f)$ è indipendente, oltre che dalla suddivisione di $[a,b]$, anche dalla scelta dei punti t_i. Una volta provato che f è integrabile, ci si può giovare di questa arbitrarietà per scegliere convenientemente la suddivisione e i punti t_i in modo da semplificare i calcoli.

***Esempio* 1.3** - Sia $f(x) = \dfrac{1}{x}, x \in [a,b], 0 < a < b$. Scegliamo gli $n+1$ punti della suddivisione nel modo seguente:

$$x_0 = a,\ x_1 = a\left(\frac{b}{a}\right)^{\frac{1}{n}},\ x_2 = a\left(\frac{b}{a}\right)^{\frac{2}{n}}, \dots, x_{n-1} = a\left(\frac{b}{a}\right)^{\frac{n-1}{n}},\ x_n = b$$

(cioè in progressione geometrica). Scegliamo poi in ogni intervallino $[x_{i-1}, x_i]$, $t_i = x_{i-1}$. Abbiamo

$$\sigma(\mathcal{D}, f) = \sum_{i=1}^{n} \frac{1}{x_{i-1}}(x_i - x_{i-1}) = \sum_{i=1}^{n}\left[\left(\frac{b}{a}\right)^{\frac{1}{n}} - 1\right] = n\left[\left(\frac{b}{a}\right)^{\frac{1}{n}} - 1\right].$$

Quando $n \to +\infty$ la massima ampiezza degli intervallini, cioè

$$\max_{1 \le i \le n}(x_i - x_{i-1}) = \max_i\left\{a\left(\frac{b}{a}\right)^{\frac{i-1}{n}}\left[\left(\frac{b}{a}\right)^{\frac{1}{n}} - 1\right]\right\} = a\left[\left(\frac{b}{a}\right)^{\frac{1}{n}} - 1\right]\left(\frac{b}{a}\right)^{\frac{n-1}{n}},$$

tende a zero; perciò possiamo calcolare l'integrale $\displaystyle\int_a^b \frac{dx}{x}$ prendendo il limite di $\sigma(\mathcal{D}, f)$ per $n \to +\infty$; otteniamo

$$\int_a^b \frac{dx}{x} = \log\frac{b}{a}.$$

La continuità non è affatto necessaria per l'integrabilità; anzi, anche funzioni con infiniti punti di discontinuità possono essere integrabili, come prova il seguente

▪ **Teorema 1.6** - *Sia $f : [a,b] \to \mathbb{R}$ limitata e monotona; allora f è integrabile.*

Dimostrazione - Sia, per fissare le idee, f crescente. Fissato $\varepsilon > 0$, per ogni suddivisione $\mathcal{D}$ con $|\mathcal{D}| < \varepsilon/(f(b) - f(a))$, risulta, ricordando il teorema 4.2.10:

$$S(\mathcal{D}, f) - s(\mathcal{D}, f) = \sum_{i=1}^{n} f(x_i)\Delta x_i - \sum_{i=1}^{n} f(x_{i-1})\Delta x_i \le$$

$$\le |\mathcal{D}| \sum_{i=1}^{n}[f(x_i) - f(x_{i-1})] = |\mathcal{D}|(f(b) - f(a)) < \varepsilon.$$

Perciò, per il teorema 1.3, f risulta integrabile. □

La condizione di monotonia indicata nel teorema 1.6 non è essenziale. Vale infatti la seguente

Proposizione 1.7 - *Sia $f : [a,b] \to \mathbb{R}$ limitata con un numero finito di punti di discontinuità; allora f è integrabile.*

Per la dimostrazione, vedere l'esercizio 1 di questa sezione.

L'affermazione contenuta nella proposizione 1.7 può essere estesa alle funzioni limitate con una infinità numerabile di punti di discontinuità (esercizio 2.12).

D'altra parte abbiamo visto che la funzione di Dirichlet su $[a,b]$, di cui l'insieme dei punti di discontinuità è tutto $[a,b]$, non è integrabile. È evidente che l'integrabiltà (secondo Riemann) di una funzione limitata dipende dall'insieme dei suoi punti di discontinuità: pur potendo questi essere infiniti, non devono però essere "troppi"; la precisazione di questo concetto richiede la nozione di *misura* di un insieme di punti sulla retta, che riteniamo prematuro introdurre a questo punto.

Osserviamo infine che una più moderna definizione di integrale, dovuta a H. Lebesgue, consente di ritenere integrabili "praticamente tutte" le funzioni limitate.

1.4 Proprietà dell'integrale

Consideriamo l'operazione di integrazione sopra definita; essa "produce" numeri a partire da funzioni. Precisamente, ad ogni funzione $f \in \mathcal{R}(I)$ ($I = [a,b]$ è un intervallo limitato) associa un ben determinato numero reale. Perciò essa è una applicazione da $\mathcal{R}(I)$ in $\mathbb{R}$, che indichiamo col simbolo $\mathcal{I}$:

$$\mathcal{I} : \mathcal{R}(I) \to \mathbb{R} \; : \; f \to \mathcal{I}(f) \, .$$

$\mathcal{R}(I)$ è il dominio di $\mathcal{I}$, $\mathbb{R}$ è il codominio.

Per questo tipo di applicazioni (corrispondenze tra un insieme di funzioni e un insieme numerico) si usa il termine più specifico di *funzionale*.

Esaminiamo ora le prime proprietà di questa applicazione.

▪ **Teorema 1.8** - *Siano* $f, g \in \mathcal{R}(I), I = [a,b]$. *Valgono allora le seguenti affermazioni(da 1. a 5.):*

1. *Additività (rispetto alla funzione integranda).* $f + g \in \mathcal{R}(I)$ *e si ha:*

$$\mathcal{I}(f+g) = \mathcal{I}(f) + \mathcal{I}(g)$$

2. *Omogeneità.* $cf \in \mathcal{R}(I)$, $\forall c \in \mathbb{R}$ *e si ha*:

$$\mathcal{I}(cf) = c\mathcal{I}(f) \, .$$

Dalle 1. e 2. discende che $\mathcal{I}$ è una *applicazione lineare.*

3. *Monotonia.*

i) $f \geq g \Rightarrow \mathcal{I}(f) \geq \mathcal{I}(g)$

ii) $|f| \in \mathcal{R}(I)$ *e si ha*:

$$|\mathcal{I}(f)| \leq \mathcal{I}(|f|)$$

In particolare, se è $M_1 = \sup_{[a,b]} |f|$, si ha

iii) $\left|\int_a^b f(x)dx\right| \leq M_1(b-a)$

La disuguaglianza *iii)* è di uso frequentissimo in Analisi.

4. *Teorema della media. Posto* $m = \inf\limits_{[a,b]} f, M = \sup\limits_{[a,b]} f$ *si ha:*

i) $m \leq \frac{1}{b-a}\int_a^b f(x)dx \leq M$

ii) Se f è continua su $[a,b]$ *esiste* $t \in [a,b]$ *tale che:*

$$\int_a^b f(x)dx = (b-a)f(t) \ .$$

L'espressione $\frac{1}{b-a}\int_a^b f(x)dx$ si chiama *valor medio* di f sull'intervallo $[a,b]$; il teorema afferma che il valor medio di f è compreso tra l'estremo inferiore e l'estremo superiore di f in $[a,b]$ e, nel caso che f sia continua, esiste (almeno) un punto in $[a,b]$ in cui f uguaglia il suo valor medio (vedi figg. 8.3a e 8.3b).

5. *Additività (rispetto all'intervallo di integrazione). Se* $c \in (a,b)$ *risulta* $f \in \mathcal{R}(a,c)$ *e* $f \in \mathcal{R}(c,b)$ *e si ha:*

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx \ . \tag{1.17}$$

Viceversa: se $f \in \mathcal{R}(a,c)$ *e* $f \in \mathcal{R}(c,b)$ *allora* $f \in \mathcal{R}(a,b)$ *e vale la* (1.17).

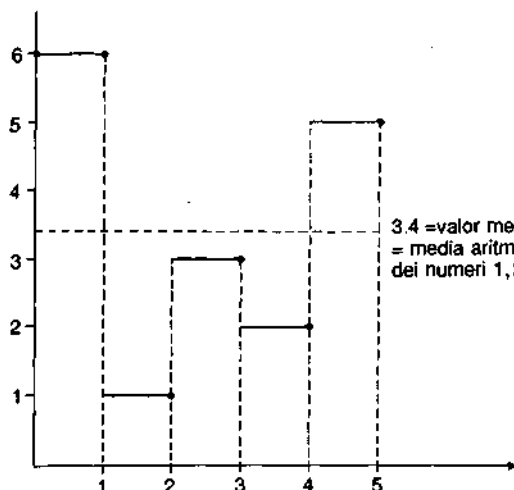
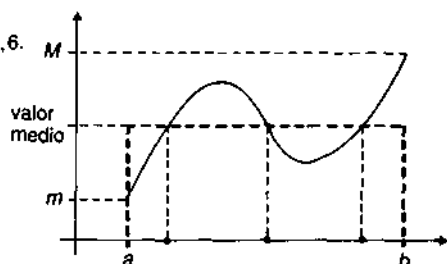



Fig. 8.3

Dimostrazione - 1. Sia $\mathcal{D}$ una qualsiasi suddivisione di I; poiché

$$\inf_{(x_{i-1},x_i)}(f+g) \geq \inf_{(x_{i-1},x_i)} f + \inf_{(x_{i-1},x_i)} g$$

$$\sup_{(x_{i-1},x_i)}(f+g) \leq \sup_{(x_{i-1},x_i)} f + \sup_{(x_{i-1},x_i)} g$$

otteniamo

$$s(\mathcal{D},f)+s(\mathcal{D},g) \leq s(\mathcal{D},f+g) \leq S(\mathcal{D},f+g) \leq S(\mathcal{D},f)+S(\mathcal{D},g)$$

e perciò

$$s(\mathcal{D},f)+s(\mathcal{D},g) \leq \sigma(\mathcal{D},f+g) \leq S(\mathcal{D},f)+S(\mathcal{D},g) \; . \tag{1.18}$$

Facendo tendere a zero l'ampiezza $|\mathcal{D}|$ il primo membro e l'ultimo della (1.18) tendono a $\mathcal{I}(f)+\mathcal{I}(g)$; perciò anche il termine centrale ammette limite (cioè $f+g \in \mathcal{R}(I)$) e tale limite $\mathcal{I}(f+g)$ coincide con $\mathcal{I}(f)+\mathcal{I}(g)$.

2. Lasciamo la dimostrazione per esercizio.

3. Osserviamo che, se $f \geq 0$, ogni somma integrale è ≥ 0 e perciò anche $\mathcal{I}(f) \geq 0$. Per provare la 3*i*) basterà applicare questa osservazione alla funzione $f-g$ e utilizzare 1. e 2.
Consideriamo ora la funzione $f_+ = \max\{f,0\}$. Se indichiamo con M_i' e m_i' rispettivamente l'estremo superiore e inferiore di f_+ nell'intervallino (x_{i-1},x_i), risulta $M_i' - m_i' \leq M_i - m_i$ $(i=1,2,\ldots,n)$; perciò si ha:

$$S(\mathcal{D},f_+) - s(\mathcal{D},f_+) = \sum_{i=1}^{n}(M_i'-m_i')\Delta x_i \leq \sum_{i=1}^{n}(M_i-m_i)\Delta x_i =$$
$$= S(\mathcal{D},f) - s(\mathcal{D},f) \; ;$$

da ciò si deduce che f_+ è integrabile. Analogamente si mostra che f_- è integrabile e perciò anche $|f| = \frac{1}{2}(f_+ + f_-)$ è integrabile. Si ha inoltre:

$$|\mathcal{I}(f)| = \left|\mathcal{I}\left(\frac{1}{2}f_+ - \frac{1}{2}f_-\right)\right| \leq \frac{1}{2}\mathcal{I}(f_+) + \frac{1}{2}\mathcal{I}(f_-) = \mathcal{I}\left(\frac{1}{2}f_+ + \frac{1}{2}f_-\right) = \mathcal{I}(|f|)$$

che prova la la 3*ii*).

Per la *iii*) basta osservare che:

$$|f(x)| \leq M_1 \text{ implica } \int_a^b |f(x)|dx \leq \int_a^b M_1 dx = M_1(b-a)$$

e perciò, applicando la *ii*), segue la tesi.

4. La 4*i*) si ottiene integrando su $[a,b]$ le disuguaglianze

$$m \leq f(x) \leq M$$

e ricordando la 3*i*) e l'esempio 1.1; si ha infatti:

$$m(b-a) \leq \int_a^b f(x)dx \leq M(b-a) \, .$$

Se f è continua su $[a,b]$ essa possiede la proprietà di Darboux, cioè assume tutti i valori compresi tra m e M; perciò in (almeno) un punto $t \in [a,b]$ assumerà anche il suo valor medio che, per quanto visto, è compreso tra m e M.

Se f è non negativa (v. fig. 8.3) il teorema afferma che esiste un rettangolo di base (a,b) e altezza compresa tra m e M che è equivalente (cioè ha la stessa area) al trapezioide.

5. Sia $\mathcal{D}$ una qualunque suddivisione di $[a,b]$ per cui, fissato $\varepsilon > 0$, risulti

$$\mathcal{S}(\mathcal{D}, f) - s(\mathcal{D}, f) < \varepsilon \, . \tag{1.19}$$

Possiamo senz'altro supporre che c sia uno dei punti della suddivisione (diversamente, consideriamo la suddivisione più fine ottenuta da $\mathcal{D}$ aggiungendo il punto c e per cui sarà ancora vera la (1.19)).

Ogni somma integrale $\sigma(\mathcal{D}, f)$ (e quindi anche la somma superiore e la somma inferiore) si può scindere in due somme: una relativa all'intervallo $[a,c]$, diciamo $\sigma_1(\mathcal{D}_1, f)$ e l'altra relativa all'intervallo $[c,b]$, diciamo $\sigma_2(\mathcal{D}_2, f)$:

$$\sigma(\mathcal{D}, f) = \sigma_1(\mathcal{D}_1, f) + \sigma_2(\mathcal{D}_2, f) \, . \tag{1.20}$$

Risulta evidentemente

$$\mathcal{S}_1(\mathcal{D}_1, f) - s_1(\mathcal{D}_1, f) \leq \mathcal{S}(\mathcal{D}, f) - s(\mathcal{D}, f) < \varepsilon$$

e analogamente per $\mathcal{S}_2(\mathcal{D}_2, f) - s_2(\mathcal{D}_2, f)$; il che prova che le restrizioni di f ad $[a,c]$ e $[c,b]$ sono integrabili. Allora, prendendo il limite per $|\mathcal{D}| \to 0$, dalla (1.20) otteniamo la (1.17).

Viceversa, siano $\mathcal{D}_1$ una suddivisione di $[a,c]$ e $\mathcal{D}_2$ una suddivisione di $[c,b]$ per cui risulti

$$\mathcal{S}_1(\mathcal{D}_1, f) - s_1(\mathcal{D}_1, f) < \varepsilon \, , \; \mathcal{S}_2(\mathcal{D}_2, f) - s_2(\mathcal{D}_2, f) < \varepsilon \, .$$

Posto $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2$, $\mathcal{D}$ è una suddivisione di $[a,b]$ per cui si ha

$$\mathcal{S}(\mathcal{D}, f) - s(\mathcal{D}, f) = \mathcal{S}_1(\mathcal{D}_1, f) + \mathcal{S}_2(\mathcal{D}_2, f) - s(\mathcal{D}_1, f) - s(\mathcal{D}_2, f) < 2\epsilon$$

per cui $f \in \mathcal{R}(a,b)$ e vale la (1.17) per quanto mostrato sopra. □

Oltre alle proprietà dell'integrale illustrate nel teorema 1.8 altre sono enunciate negli Esercizi 2, 3, 4, 5, 6.

Esaminiamo ora il comportamento dell'integrale sopra definito in relazione all'orientamento dell'intervallo di integrazione. Nel paragrafo 2.2.4, definendo gli intervalli $[a,b]$, (a,b), etc. si era assunto $a < b$; su tali intervalli risulta, in modo naturale, definito un orientamento, da a verso b e si diranno perciò *orientati positivamente*.

Conveniamo ora che la scrittura $[a,b]$ (e similmente $[a,b)\ldots$) con $a > b$ indichi ancora un intervallo che, dal punto di vista insiemistico, coincide con $[a,b]$, ma che ha *orientamento negativo*.

L'integrale è stato sopra definito per intervalli orientati positivamente. Conviene ora estendere la definizione anche agli intervalli orientati negativamente, ponendo, se $b < a$:

$$\int_a^b f(x)dx := -\int_b^a f(x)dx.$$

Porremo anche, nel caso $a = b$:

$$\int_a^a f(x)dx := 0\ .$$

Si riconosce facilmente che le proprietà 1., 2., 4. del teorema 1.8 continuano a valere indipendentemente dall'orientamento dell'intervallo di integrazione; le proprietà di monotonia 3. si modificano in maniera evidente; la proprietà 5. si estende nella maniera seguente:

▪ **Teorema 1.9** - *Sia $f \in \mathcal{R}(I)$; comunque si prendano $a,b,c \in I$ vale la relazione:*

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx\ . \tag{1.21}$$

Dimostrazione - Basta esaminare, uno per uno, i vari casi possibili in relazione all'ordinamento dei tre numeri a,b,c. Se $a < c < b$ la (1.21) coincide con la (1.18).

Se $c < a < b$ abbiamo, per la proprietà 5. del teorema 1.8:

$$\int_c^b f(x)dx = \int_c^a f(x)dx + \int_a^b f(x)dx$$

da cui otteniamo

$$\int_a^b f(x)dx = \int_c^b f(x)dx - \int_c^a f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx$$

e così via per gli altri casi. □

1.5 Il teorema fondamentale del calcolo integrale. Integrale indefinito

L'integrale precedentemente introdotto, cioè l'integrale di una funzione f esteso ad un determinato intervallo, viene detto **integrale definito**.

Sia ora $f \in \mathcal{R}(a,b)$ e sia $c \in [a,b]$; se x è un altro punto qualsiasi di $[a,b]$ la funzione f è certo integrabile sull'intervallo $[c,x]$; poniamo

$$F(x) := \int_c^x f(t)dt \tag{1.22}$$

La (1.22) chiaramente definisce una funzione $F : [a,b] \to \mathbb{R}$, che chiameremo **funzione integrale** di f (relativa al punto c).

▪ **Teorema 1.10** - (teorema fondamentale del Calcolo integrale). *Sia $f \in C^0([a,b])$ e sia $c \in [a,b]$; allora la funzione integrale F* (definita dalla (1.22)) *è di classe $C^1([a,b])$ e si ha : $F'(x) = f(x)\ \forall x \in [a,b]$.*

Dimostrazione - Sia $x \in [a,b]$ e h un incremento abbastanza piccolo in modo che anche $x+h \in [a,b]$; si ha (usando la (1.21)):

$$\begin{aligned} F(x+h) - F(x) &= \int_c^{x+h} f(t)dt - \int_c^x f(t)dt = \\ &= \int_x^c f(t)dt + \int_c^{x+h} f(t)dt = \\ &= \int_x^{x+h} f(t)dt = hf(\xi) \end{aligned}$$

essendo ξ un punto dell'intervallo $[x, x+h]$: nell'ultimo passaggio si è fatto uso della proprietà della media (4 *ii*) del teorema 1.8). Osserviamo ora che, quando $h \to 0$, ξ tende al punto x; allora $f(\xi)$ (essendo f continua) tende a $f(x)$; abbiamo allora:

$$\lim_{h\to 0} \frac{F(x+h) - F(x)}{h} = \lim_{h\to 0} f(\xi) = f(x)$$

che prova la tesi. □

Il teorema 1.10 mostra che *ogni funzione integrale di una funzione continua f è una* **primitiva** *di f*; poiché sappiamo (vedi 6.2.3) che due primitive di una stessa funzione differiscono tra loro per una costante, ne segue che *l'insieme di tutte le primitive di una funzione f continua su I è dato dall'espressione*

$$\int_c^x f(t)dt + k \tag{1.23}$$

dove c è un punto fissato di I e k è una costante arbitraria.

Sottolineiamo che sotto la sola ipotesi di integrabilità, non si può affermare che F è derivabile; ad esempio, se $f(x) = \operatorname{sgn} x$, allora $F(x) = \int_0^x f(t)dt = |x|$, non derivabile nell'origine. Osserviamo tuttavia che F è continua (v. esercizio 8).

La famiglia di tutte le primitive di una funzione assegnata prende il nome di **integrale indefinito.** Il simbolo $\int f(x)dx$ viene indifferentemente usato per indicare la famiglia di tutte le primitive di $f(x)$ oppure una qualsiasi di esse.

Le precedenti osservazioni giustificano l'affermazione, impropria ma suggestiva: l'integrazione è l'operazione inversa della derivazione. (Lo studente noterà che la derivazione, essendo un'operazione non iniettiva (infatti se f è derivabile, f e $f+$ costante hanno la stessa derivata!) non è invertibile).

Il seguente corollario del teorema 1.10 è della massima utilità.

Corollario 1.11 - *Sia f continua su $[a,b]$ e sia G una primitiva di f. Allora risulta*

$$\int_a^b f(x)dx = G(b) - G(a)\,. \tag{1.24}$$

Dimostrazione - Se G è una primitiva di f, esiste una costante k per cui risulta

$$G(x) = \int_a^x f(t)dt + k$$

(si è preso $c = a$ nella (1.23)). Ponendo, nella formula precedente, $x = a$ si trova $G(a) = k$; si ha perciò

$$G(x) = \int_a^x f(t)dt + G(a)$$

Ponendo in questa $x = b$ si ottiene la (1.24). □

Basterà dunque conoscere una primitiva di f (funzione continua su $[a,b]$) per poter calcolare immediatamente l'integrale definito di f tra a e b.

L'espressione $G(b) - G(a)$, cioè l'incremento di G nel passare da a a b, viene solitamente indicato con una delle notazioni seguenti:

$$[G(x)]_a^b \quad , \quad G(x)\,\Big|_a^b \ .$$

***Esempio* 1.4** - Poiché $\log x$, $0 < a \le x \le b$, è una primitiva di $1/x$ in $[a,b]$ si trova subito

$$\int_a^b \frac{1}{x}dx = \log b - \log a$$

come avevamo già trovato nell'esempio 1.3, con un po' di calcoli.

Dalla tabella di derivate in 6.1.4 si ricava subito la tabella di primitive riportata a pagina seguente.

Come abbiamo già osservato in 6.1.4 tutte le funzioni elementari sono derivabili (in ogni punto dell'insieme di definizione) e la derivata è ancora una funzione elementare; inoltre, essendo continue, le funzioni elementari ammettono primitive. Ci si può chiedere se anche le primitive delle funzioni elementari sono ancora funzioni

elementari; la risposta è negativa, come vedremo; una funzione che ammette come primitiva una funzione elementare si dirà *integrabile elementarmente.*

Il primo risultato importante in questa direzione è contenuto nella seguente

Proposizione 1.12 - *Se $f(x)$ è una funzione razionale essa è integrabile elementarmente; ogni primitiva è una combinazione lineare di funzioni razionali e di funzioni del tipo:* $\log(ax^2 + bx + c)$ *e* $\operatorname{arctg}(\alpha x + \beta)$.

La proposizione è dimostrata e illustrata con esempi nell'esercizio 1.8.

Passando alla classe delle funzioni algebriche incontriamo i primi esempi di funzioni non integrabili elementarmente.

Funzione	Una primitiva
$x^\alpha \quad (\alpha \neq -1)$	$\dfrac{x^{\alpha+1}}{\alpha+1}$
$\lvert x\rvert^\alpha \quad (\alpha \neq -1)$	$\dfrac{x\lvert x\rvert^\alpha}{\alpha+1}$
$\dfrac{1}{x}$	$\log\lvert x\rvert$
$\sin x$	$-\cos x$
$\cos x$	$\sin x$
$\dfrac{1}{(\cos x)^2}$	$\operatorname{tg} x$
$\dfrac{1}{(\sin x)^2}$	$-\operatorname{cotg} x$
a^x	$\dfrac{a^x}{\log a}$
e^x	e^x
$\operatorname{Sh} x$	$\operatorname{Ch} x$
$\operatorname{Ch} x$	$\operatorname{Sh} x$
$\dfrac{1}{(\operatorname{Ch} x)^2}$	$\operatorname{Th} x$
$\dfrac{1}{1+x^2}$	$\operatorname{arctg} x$
$\dfrac{1}{\sqrt{1-x^2}}$	$\arcsin x$

Sia $y = y(x)$ un *ramo* della *curva algebrica* Γ definita dall'equazione $P(x, y) = 0$ dove P è un polinomio nelle indeterminate x e y. Si chiama *integrale abe-*

liano (*) (appartenente alla curva algebrica Γ) un integrale del tipo

$$\int R(x, y(x))dx \tag{1.25}$$

dove $R(x, y)$ è una funzione razionale in x e y.
Per esempio, se Γ è la *conica* di equazione

$$y^2 - ax^2 - bx - c = 0$$

abbiamo integrali della forma

$$\int R(x, \sqrt{ax^2 + bx + c})dx \ . \tag{1.26}$$

Questi integrali sono tutti calcolabili elementarmente (vedi esercizio 20).
Invece integrali abeliani che si presentano sotto la forma

$$\int R\left(x, \sqrt{Q(x)}\right)dx \tag{1.27}$$

dove Q è un polinomio di grado superiore al secondo, non sono in generale calcolabili elementarmente. Essi "definiscono" pertanto nuove funzioni; per esempio, se Q è un polinomio di terzo o quarto grado, gli integrali (1.27) conducono alla definizione delle cosiddette *funzioni ellittiche*, che intervengono nel problema della rettificazione dell'ellisse e di altre curve algebriche come le lemniscate.

Passando poi alle funzioni elementari trascendenti, si incontrano moltissimi esempi di integrali non esprimibili per mezzo di funzioni elementari e che hanno una rilevante importanza in svariati problemi di Matematica, di Fisica, di Ingegneria.

Le funzioni (non elementari) definite per mezzo di integrali vengono dette trascendenti non elementari, o, più comunemente, *funzioni speciali*. Lo studio di queste funzioni esula completamente dai limiti di questa trattazione.
In questo ordine di idee risulta utile il seguente teorema, che ci limitiamo ad enunciare.

■ **Teorema 1.13** - (di Liouville). *Siano $g(x)$ ed $h(x)$ funzioni razionali (anche a valori complessi) con $g(x)$ non costante. Allora, se*

$$\int e^{g(x)}h(x)dx$$

è una funzione elementare, essa è della forma

$$e^{g(x)}r(x) + \ \textit{costante}$$

(*) Dal nome del matematico norvegese N.H. Abel (1802–1828).

dove $r(x)$ è una funzione razionale.

Dal punto di vista pratico, il teorema si utilizza nel modo seguente. Sia $r(x) = p(x)/q(x)$, con p e q polinomi primi tra loro. Se $\exp(g(x))\, p(x)/q(x)$ è una primitiva di $\exp(g(x))h(x)$, deve essere

$$\frac{d}{dx}\left[e^{g(x)}\frac{p(x)}{q(x)}\right] = e^{g(x)}h(x)$$

e cioè, eseguendo i calcoli,

$$e^{g(x)}\left\{\frac{p(x)}{q(x)}g'(x) + \frac{p'(x)q(x) - p(x)q'(x)}{q(x)^2}\right\} = e^{g(x)}h(x) .$$

Fatte alcune semplificazioni, si ottiene

$$\{p'(x) + g'(x)p(x) - h(x)q(x)\}q(x) = p(x)q'(x) . \tag{1.28}$$

Pertanto $e^{g(x)}h(x)$ *ha primitiva elementare se e solo se esistono due polinomi p e q, primi tra loro, verificanti la* (1.28).

Per esempio, *la funzione e^{ax^2} $(a \in \mathbb{C}, a \neq 0)$ non ha primitiva elementare.* Infatti la (1.28) si scrive

$$(p'(x) + 2axp(x) - q(x))q(x) = p(x)q'(x) . \tag{1.29}$$

Questa relazione è impossibile. Infatti, sia λ uno zero di q di molteplicità m; avremo $q(x) = (x-\lambda)^m t(x)$ con $t(\lambda) \neq 0$ e $m \geq 1$; $q'(x) = (x-\lambda)^{m-1}[mt(x) + (x-\lambda)t'(x)]$, perciò λ è uno zero di molteplicità $m-1$ per il secondo membro della (1.29) (ricorda che λ non può essere zero di p poiché p e q sono primi tra loro), mentre è uno zero di molteplicità m o maggiore per il primo membro; assurdo. Dobbiamo ammettere allora che q sia costante e la (1.29) diventa: $p' + 2axp = q$; ma anche questa relazione è assurda poiché il primo membro è un polinomio di grado ≥ 1, mentre il secondo membro è costante.

Una primitiva di e^{-x^2} è stata estesamente studiata da Gauss, in relazione a problemi del Calcolo delle Probabilità. Più precisamente, si definisce la *funzione degli errori* ponendo (*)

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$$

Tale funzione è tabulata su molti testi di Statistica.

Non indugiamo oltre sul problema di trovare le primitive delle funzioni elementari. Dal punto di vista pratico ogni studente può facilmente procurarsi delle

(*) erf sta per: erroris functio.

estese tavole di integrali calcolabili elementarmente, sufficienti per le più comuni applicazioni. Ricordiamo anche che oggi si può effettuare l'integrazione simbolica ed algebrica direttamente su un calcolatore.

1.6 Regole di integrazione

La precedente tabella di primitive può essere considerevolmente ampliata con l'aiuto di alcune semplici regole (o metodi) di integrazione, che ora illustriamo.

Integrazione per scomposizione.

Questo metodo è basato sulle proprietà 1. e 2. del teorema 1.8. Si cerca di esprimere (decomporre) la funzione integranda come combinazione lineare di più funzioni integrabili immediatamente.

***Esempio* 1.5** - Si voglia calcolare

$$\int \frac{dx}{(\cos x)^2 (\sin x)^2}$$

Possiamo scrivere

$$\frac{1}{(\cos x)^2 (\sin x)^2} = \frac{1}{(\cos x)^2} + \frac{1}{(\sin x)^2} \tag{!}$$

cosicché abbiamo

$$\int \frac{dx}{(\cos x)^2 (\sin x)^2} = \int \frac{dx}{(\cos x)^2} + \int \frac{dx}{(\sin x)^2} = \operatorname{tg} x - \frac{1}{\operatorname{tg} x} + \text{ costante} .$$

Integrazione per parti.

Questo metodo è basato sulla regola di derivazione del prodotto di due funzioni:

$$(f(x)g(x))' = f'(x)g(x) + f(x)g'(x) . \tag{1.30}$$

Se f e g sono continue con le derivate f' e g' in un intervallo $[a, b]$, dalla (1.30), integrando su $[a, b]$, abbiamo:

$$[f(x)g(x)]_a^b = \int_a^b f'(x)g(x)dx + \int_a^b f(x)g'(x)dx$$

che viene utilizzata nella forma seguente

$$\int_a^b f(x)g'(x)dx = -\int_a^b f'(x)g(x)dx + [f(x)g(x)]_a^b . \tag{1.31}$$

Nell'integrale a sinistra della (1.31) il termine $f(x)$ prende il nome di *fattore finito*, mentre $g'(x)dx$ vien detto *fattore differenziale*.

Integrando la (1.30) su $[a,x]$, con $x \in [a,b]$, si ottiene la seguente relazione per gli integrali indefiniti:

$$\int f(x)g'(x)dx = -\int f'(x)g(x)dx + f(x)g(x) . \tag{1.32}$$

***Esempio* 1.6** - Si voglia calcolare $\int \log x dx$. Si può utilizzare la (1.32) prendendo come fattore finito $f(x) = \log x$ e come fattore differenziale $1 \cdot dx$ (cioè $g(x) = x$). Otteniamo:

$$\int \log x dx = -\int \frac{1}{x} x dx + x \log x = -x + x \log x + \text{ costante } .$$

Integrazione per sostituzione.

Questo metodo è basato sulla regola di derivazione di funzione composta.

Sia $f : [a,b] \to \mathbb{R}$ continua e $\varphi : [\alpha,\beta] \to [a,b]$ di classe C^1 (cioè continua e derivabile con continuità).

Supponiamo inoltre che $\varphi([\alpha,\beta]) = [a,b]$ cioè, al variare di t in $[\alpha,\beta], \varphi(t)$ descriva l'intervallo $[a,b]$. Preso allora un qualunque intervallo $[c,d] \subseteq [a,b]$ esistono due valori (almeno) γ e δ tali che risulta

$$c = \varphi(\gamma) \qquad d = \varphi(\delta) . \tag{1.33}$$

Sia ora $F(x)$ una primitiva di $f(x)$ in $[a,b]$; abbiamo allora:

$$\int_c^d f(x)dx = F(d) - F(c) .$$

Per la regola di derivazione di funzione composta, si ha

$$(F \circ \varphi)'(t) = F'(\varphi(t))\varphi'(t) = f(\varphi(t))\varphi'(t) \tag{1.34}$$

cioè $F \circ \varphi(t)$ è una primitiva di $f(\varphi(t))\varphi'(t)$ in $[\alpha,\beta]$. Abbiamo allora, integrando la (1.34) su $[\gamma,\delta]$:

$$F(\varphi(\delta)) - F(\varphi(\gamma)) = \int_\gamma^\delta f(\varphi(t))\varphi'(t)dt$$

da cui, per le (1.33)

$$\int_c^d f(x)dx = \int_\gamma^\delta f(\varphi(t))\varphi'(t)dt . \tag{1.35}$$

Si osservi che l'intervallo $[\gamma,\delta]$ non è, in generale, univocamente definito; lo è se φ è invertibile, cioè, per esempio, se $\varphi'(t) \neq 0$ in (α,β). In tal caso si potrà scrivere

$$\int_c^d f(x)dx = \int_{\varphi^{-1}(c)}^{\varphi^{-1}(d)} f(\varphi(t))\varphi'(t)dt \,. \tag{1.36}$$

Per l'integrale indefinito varrà allora la formula (che si ottiene dalla (1.34) integrando su $[c,x]$ con $x \in [a,b]$, $x = \varphi(t)$

$$\int f(x)dx = \int f(\varphi(t))\varphi'(t)dt \,. \tag{1.37}$$

Si noti che, *formalmente*, nella (1.37) si passa da un integrale all'altro ponendo $x = \varphi(t)$ e, "conseguentemente", $dx = \varphi'(t)dt$; ciò indica quanto sia appropriata la notazione utilizzata per indicare l'integrale.

***Esempio* 1.7** - Si voglia calcolare $\int_0^1 e^{\sqrt{x}}dx$. Poniamo $x = t^2$ e utilizziamo la (1.35) con $\varphi(t) = t^2$ e $[c,d] = [0,1]$. Abbiamo due scelte per l'intevallo $[\gamma,\delta]$: possiamo prendere $[0,1]$ oppure $[0,-1]$. Nel primo caso abbiamo (da $x = t^2$ ricaviamo $\sqrt{x} = t$)

$$\int_0^1 e^{\sqrt{x}}dx = \int_0^1 e^t \cdot 2t dt = \text{(integrando per parti)}$$

$$= 2\left[te^t \Big|_0^1 - \int_0^1 e^t dt\right] = 2\left[e - e^t \Big|_0^1\right] = 2 \,.$$

Scegliendo per $[\gamma,\delta]$ l'intervallo $[0,-1]$ avremmo (ora da $x = t^2$ ricaviamo $\sqrt{x} = -t$)

$$\int_0^1 e^{\sqrt{x}}dx = \int_0^{-1} e^{-t} \cdot 2t dt$$

che, naturalmente, porta allo stesso risultato.

Osservazione 1.3 - (Cambi di variabile come cambi di scala).

Il cambio di variabile $x = at$ è interpretabile come cambio di scala (unità di misura) sull'asse delle ascisse. Ad esempio, se x è misurato in metri e $a = 100$, t risulta misurato in centimetri. Così, nelle somme superiori ed inferiori una lunghezza di Δx metri viene ora misurata in $100\Delta t$ centimetri. Se x varia tra 0 e 1 metri, t varia da 0 a 100 cm; questo si riflette nell'integrale dove dx (simbolo per la misura usata) si trasforma in $100dt$.

Nel caso di cambiamenti lineari il cambio di scala è identico in ogni punto dell'intervallo di integrazione ed è realizzato tramite il fattore di conversione a.

Se si pone $x = \varphi(t)$ con $\varphi(t)$ non lineare, la formula (1.35) indica che dx si è trasformato nel differenziale $d\varphi = \varphi'(t)dt$. Ciò si può interpretare come un cambiamento di scala in cui il fattore di conversione $\varphi'(t)$ varia da punto a punto nell'intervallo di integrazione.

1.7 Integrazione numerica

Sia $f \in \mathcal{R}(a,b)$ e si voglia calcolare $\int_a^b f(x)dx$. Quando l'integrale non sia calcolabile elementarmente o, comunque, non sia agevole procurarsi una primitiva di $f(x)$, si può ricorrere a formule di integrazione numerica. Esse consistono nel sostituire alla funzione integranda $f(x)$ una sua approssimazione $\varphi(x)$ per la quale il calcolo dell'integrale risulti semplice e rapido.

Per esempio (e questa è senz'altro la prima idea che viene in mente, suggerita dalla definizione stessa dell'integrale) si può assumere per $\varphi(x)$ una funzione *costante a pezzi*. Più precisamente, prendiamo una suddivisione $\mathcal{D}(x_0, \ldots, x_n)$ dell'intervallo $[a,b]$ con punti equidistanti, cioè

$$x_i = a + \frac{b-a}{n} i \qquad (i = 0, 1, \ldots, n)$$

e poniamo

$$\varphi(x) = f(x_i) \quad \text{per } x_i \leq x < x_{i+1} \quad (i = 0, 1, \ldots, n-1) .$$

Risulta allora

$$S_n = \int_a^b \varphi(x)dx = \frac{b-a}{n}(f(x_0) + f(x_1) + \ldots + f(x_{n-1})) . \tag{1.38}$$

Il teorema 1.4 ci autorizza ad assumere S_n come valore approssimato dell'integrale $\mathcal{I} = \int_a^b f(x)dx$, nel senso che, per $n \to +\infty$, S_n tende al valore vero $\mathcal{I}$.

Perché un tale procedimento sia utilizzabile nelle applicazioni bisogna però conoscere almeno una stima dell'errore che si commette sostituendo $\mathcal{I}$ con S_n; cioè, oltre alla formula di integrazione (1.38) serve anche una formula per la maggiorazione dell'errore. Questo è uno dei compiti principali dell'Analisi numerica.

Per esempio, nel caso della (1.38) si può dimostrare che:

se f è continua e derivabile in $[a,b]$ e se $|f'(x)| \leq M$, $\forall x \in [a,b]$, allora

$$|\mathcal{I} - S_n| \leq \frac{M}{2}\frac{(b-a)^2}{2n} . \tag{1.39}$$

Il procedimento sopra illustrato (detto *metodo dei rettangoli*: vedi fig. 8.4*a*)), non è molto conveniente per le applicazioni, poiché, come mostra la (1.39), occorre

prendere n molto grande per avere buone approssimazioni. Procedimenti più convenienti sono: il *metodo delle tangenti* (fig. 8.4*b*): in ogni intervallino $[x_i, x_{i+1}]$ il grafico di $f(x)$ viene sostituito con la tangente a detto grafico nel punto di mezzo; il *metodo dei trapezi* (fig. 8.4*c*): in ogni intervallino si assume come grafico di $\varphi(x)$ il segmento di retta congiungente i punti $(x_i, f(x_i))$ e $(x_{i+1}, f(x_{i+1}))$.

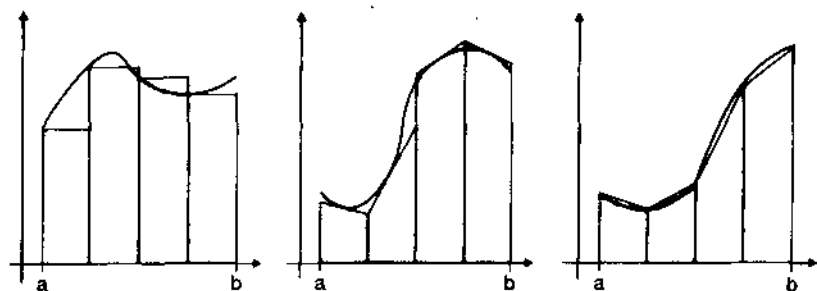



Fig. 8.4 *a)* Metodo dei rettangoli *b)* Metodo delle tangenti *c)* Metodo dei trapezi

Uno dei metodi più soddisfacenti, e quindi di più largo impiego, è il *metodo di Cavalieri–Simpson*. Esso consiste nell'approssimare la funzione integranda, in ogni intervallino della suddivisione, invece che con segmenti di rette, con archi di parabole. Precisamente, data la suddivisione $\mathcal{D}(x_0, x_1, \dots, x_n)$, indichiamo con z_i $(i = 1, 2, \dots, n)$ il punto medio dell'intervallino $[x_{i-1}, x_i]$.

Consideriamo ora la parabola, con l'asse parallelo all'asse y, e passante per i punti $(x_{i-1}, f(x_{i-1}))$, $(z_i, f(z_i))$, $(x_i, f(x_i))$. La sua equazione potrà scriversi nella forma

$$\varphi(x) = \alpha(x - z_i)^2 + \beta(x - z_i) + \gamma \qquad x_{i-1} \leq x \leq x_i$$

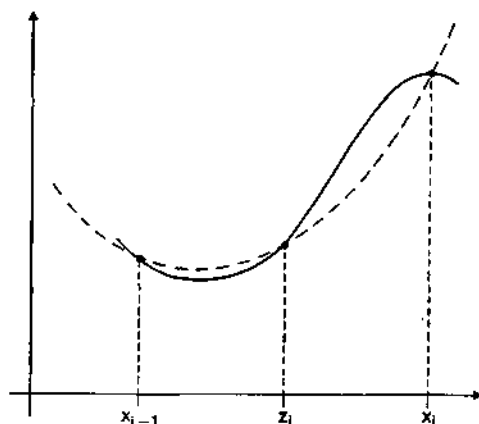



Fig. 8.5

dove le costanti α, β, γ saranno determinate dalle condizioni:

$$\begin{aligned} f(x_{i-1}) &= \alpha(x_{i-1} - z_i)^2 + \beta(x_{i-1} - z_i) + \gamma \\ f(z_i) &= \gamma \\ f(x_i) &= \alpha(x_i - z_i)^2 + \beta(x_i - z_i) + \gamma \ . \end{aligned}$$

Utilizzando, come quasi sempre si fa in pratica, una suddivisione in intervalli uguali, ciascuno di ampiezza $2\delta = (b-a)/n$, si ha

$$\begin{aligned} f(x_{i-1}) - f(z_i) &= \alpha\delta^2 - \beta\delta \\ f(x_i) - f(z_i) &= \alpha\delta^2 + \beta\delta \ . \end{aligned} \tag{1.40}$$

Si ottiene allora

$$\begin{aligned} \int_{x_{i-1}}^{x_i} \varphi(x)dx &= \int_{x_{i-1}}^{x_i} [\alpha(x - z_i)^2 + \beta(x - z_i) + \gamma]dx = \\ &= \left[\alpha\frac{(x - z_i)^3}{3} + \beta\frac{(x - z_i)^2}{2} + \gamma x\right]_{x_{i-1}}^{x_i} = \\ &= \frac{2}{3}\alpha\delta^3 + 2\delta\gamma \ . \end{aligned}$$

Ricavando α dalle (1.40) e tenendo conto che $\gamma = f(z_i)$ abbiamo

$$\int_{x_{i-1}}^{x_i} \varphi(x)dx = \frac{b-a}{n} \ \frac{f(x_i) + f(x_{i-1}) + 4f(z_i)}{6} \ .$$

Sommando su i da 1 a n otteniamo la formula (di Cavalieri–Simpson):

$$\begin{aligned} S'_n = \int_a^b \varphi(x)dx = \frac{b-a}{bn}\{&f(x_0) + f(x_n) + \\ &+ 2[f(x_1) + f(x_2) + \ldots + f(x_{n-1})] + \\ &+ 4[f(z_1) + f(z_2) + \ldots + f(z_n)]\} \ . \end{aligned} \tag{1.41}$$

Anche per la formula (1.41) si ha una stima dell'errore. Si può mostrare infatti che: *se* $f \in C^4(a,b)$ *e se* $|f^{(4)}(x)| \leq M'$ *in* $[a,b]$, *allora risulta*

$$|\mathcal{I} - S'_n| \leq \frac{M'}{2880} \ \frac{(b-a)^5}{n^4} \ .$$

Per esercizio, l'allievo potrà ricavare le formule di integrazione relative al metodo delle tangenti e al metodo dei trapezi.

1.8 Integrali dipendenti da un parametro

Supponiamo che la funzione limitata $x \mapsto f(x,y)$ sia integrabile su $[a,b]$ per ogni $y \in [c,d]$; l'integrale $\int_a^b f(x,y)dx$ è una funzione di y, diciamo $\Phi(y)$:

$$\Phi(y) = \int_a^b f(x,y)dx \; . \tag{1.42}$$

Ci domandiamo come si comporta l'integrale rispetto alle operazioni "infinitesimali": passaggio al limite e derivazione; più esplicitamente, volendo calcolare

$$\lim_{y \to y_0} \Phi(y) \quad \text{e} \quad \frac{d}{dy}\Phi(y)$$

è possibile procedere invertendo l'ordine delle operazioni e calcolare

$$\int_a^b \lim_{y \to y_0} f(x,y)dx \quad \text{e} \quad \int_a^b \frac{\partial}{\partial y} f(x,y)dx \; ?$$

Daremo, in questo paragrafo, una prima risposta a questi quesiti, che saranno poi ripresi nel Vol. II.

Sia $x \in I = [a,b]$ e $y \in J = [c,d]$. Poniamo $E = I \times J$ e consideriamo la funzione $f : E \to \mathbb{R}$.

Proposizione 1.14 - *Se f è continua in E, allora la funzione Φ, definita dalla* (1.42), *è continua in J. In particolare, se $y_0 \in J$, abbiamo*

$$\lim_{y \to y_0} \Phi(y) = \int_a^b \lim_{y \to y_0} f(x,y)dx \tag{1.43}$$

(formula di passaggio al limite sotto il segno di integrale).

La proposizione 1.14 è un semplice corollario del seguente teorema, nel quale si considera anche la dipendenza dell'integrale dagli estremi di integrazione.
Siano dati $\alpha \in [a,b], \beta \in [a,b], y \in [c,d]$ e consideriamo l'integrale

$$\Phi(\alpha,\beta,y) = \int_\alpha^\beta f(x,y)dx \; . \tag{1.44}$$

La funzione Φ è ora definita su $H = I \times I \times J$.

▪ **Teorema 1.15** - *Se f è continua in E, la funzione Φ, definita dalla* (1.44), *è continua in H.*

Dimostrazione - Sia $(\alpha_0, \beta_0, y_0) \in H$; fissato $\varepsilon > 0$ abbiamo

$$|\Phi(\alpha,\beta,y) - \Phi(\alpha_0,\beta_0,y_0)| = \left|\int_\alpha^\beta f(x,y)dx - \int_{\alpha_0}^{\beta_0} f(x,y_0)dx\right|$$

$$= \left|\int_\alpha^{\alpha_0} f(x,y)dx + \int_{\alpha_0}^{\beta_0}[f(x,y) - f(x,y_0)]dx + \int_{\beta_0}^{\beta} f(x,y)dx\right| \leq$$

$$\leq \int_\alpha^{\alpha_0} |f(x,y)|dx + \int_{\alpha_0}^{\beta_0} |f(x,y) - f(x,y_0)|dx + \int_{\beta_0}^{\beta} |f(x,y)|dx \leq$$

(utilizzando la $3iii$) del teorema 1.8, con $M = \max_E |f|$)

$$\leq |\alpha_0 - \alpha|M + \int_{\alpha_0}^{\beta_0} |f(x,y) - f(x,y_0)|dx + |\beta - \beta_0|M \ .$$

Poiché f è continua sull'insieme compatto E, è uniformemente continua e perciò si può trovare $\delta = \delta(\varepsilon) > 0$ tale che si abbia

$$|f(x,y) - f(x,y_0)| < \varepsilon$$

se

$$|(x,y) - (x,y_0)| = |y - y_0| < \delta \ .$$

Preso dunque y tale che $|y - y_0| < \delta$ e α e β tali che $|\alpha - \alpha_0| < \varepsilon$ e $|\beta - \beta_0| < \varepsilon$ risulta

$$|\Phi(\alpha,\beta,y) - \Phi(\alpha_0,\beta_0,y_0)| < (2M + |\beta_0 - \alpha_0|)\varepsilon \ .$$

Per l'arbitrarietà di ε, concludiamo che Φ è continua in (α_0, β_0, y_0); per l'arbitrarietà di (α_0, β_0, y_0) segue la tesi. □

Veniamo ora al secondo problema.

▪ **Teorema 1.16** - *Sia f continua in E, insieme con la derivata parziale f_y. Allora la funzione Φ, definita dalla* (1.43), *è derivabile e si ha*

$$\frac{d\Phi}{dy}(y) = \int_a^b \frac{\partial}{\partial y} f(x,y)dx \tag{1.45}$$

(formula di derivazione sotto il segno di integrale).

Dimostrazione - Preso un punto $y \in [c,d]$ e un incremento h abbastanza piccolo, si ha

$$\frac{\Phi(y+h) - \Phi(y)}{h} = \int_a^b \frac{f(x,y+h) - f(x,y)}{h}dx =$$

$$= \int_a^b f_y(x,y)dx + \int_a^b \left[\frac{f(x,y+h) - f(x,y)}{h} - f_y(x,y)\right]dx =$$

(per il teorema del valor medio di Lagrange)

$$= \int_a^b f_y(x,y)dx + \int_a^b [f_y(x,y+\theta h) - f_y(x,y)]dx$$

con $\theta = \theta(x, y, h)$, $0 < \theta < 1$.

Poiché, per l'ipotesi, f_y è uniformemente continua in E, abbiamo che, procedendo come nella dimostrazione del teorema 1.12, fissato $\varepsilon > 0$, esiste $\delta > 0$ tale che, se $|h| < \delta$,

$$|f_y(x, y + \theta h) - f_y(x, y)| < \varepsilon .$$

Allora abbiamo

$$\left|\frac{\Phi(y+h) - \Phi(y)}{h} - \int_a^b f_y(x, y)dx\right| \leq \int_a^b |f_y(x, y+\theta h) - f_y(x, y)|dx \leq \varepsilon(b-a) ,$$

che prova la tesi. □

Anche il teorema 1.16 può generalizzarsi al caso in cui gli estremi di integrazione dipendano a loro volta da y. Consideriamo cioè integrali della forma

$$\Psi(y) = \int_{\alpha(y)}^{\beta(y)} f(x, y)dx \tag{1.46}$$

nell'ipotesi che:

$$y \in J = [c, d] \ , \ \alpha(y) \in I = [a, b] \ , \ \beta(y) \in I = [a, b] .$$

Allora Ψ è definita su J.

Corollario 1.17 - *Se f è continua in $E = I \times J$ insieme con la sua derivata parziale f_y, e se α e β sono di classe $C^1(J)$, allora Ψ è di classe $C^1(J)$ e risulta*

$$\Psi'(y) = \int_{\alpha(y)}^{\beta(y)} f_y(x, y)dx + \beta'(y)f(\beta(y), y) - \alpha'(y)f(\alpha(y), y) . \tag{1.47}$$

Dimostrazione - Risulta infatti, se Φ è la funzione definita dalla (1.44),

$$\Psi(y) = \Phi(\alpha(y), \beta(y), y)$$

e perciò, per la regola di derivazione di funzione composta,

$$\Psi'(y) = \frac{\partial\Phi}{\partial\alpha}\alpha'(y) + \frac{\partial\Phi}{\partial\beta}\beta'(y) + \frac{\partial\Phi}{\partial y} .$$

Dal teorema fondamentale del calcolo integrale abbiamo

$$\frac{\partial}{\partial\beta}\Phi(\alpha, \beta, y) = \frac{\partial}{\partial\beta}\int_\alpha^\beta f(x, y)dx = f(\beta, y)$$

$$\frac{\partial\Phi}{\partial\alpha}(\alpha, \beta, y) = \frac{\partial}{\partial\alpha}\int_\alpha^\beta f(x, y)dx = \frac{\partial}{\partial\alpha}\left[-\int_\beta^\alpha f(x, y)dx\right] = -f(\alpha, y) .$$

Tenuto conto anche della (1.45) si ricava la (1.47). □

Come lo studente avrà notato, i teoremi di questo paragrafo utilizzano, in maniera sostanziale, la uniforme continuità della funzione integranda. Essi inoltre si estendono senza difficoltà al caso in cui $\mathbf{y}$ sia un elemento di $\mathbb{R}^n$, cioè ad integrali della forma

$$\int_{\alpha(\mathbf{y})}^{\beta(\mathbf{y})} f(x, \mathbf{y}) dx \ .$$

Esempi

1.8. La funzione $f(x,y) = \dfrac{\exp(x^2 y)}{x^2 + y^2 + 1}$ è continua in $\mathbb{R}^2$ (e perciò uniformemente continua in ogni insieme limitato). Si ha pertanto

$$\lim_{y \to 0} \int_{-1}^{+1} \frac{\exp(x^2 y)}{x^2 + y^2 + 1} dx = \int_{-1}^{+1} \frac{dx}{1 + x^2} = \frac{\pi}{2} \ .$$

1.9. Si consideri la funzione

$$\Phi(y) = \int_1^{\sqrt{y}} \frac{e^{-xy}}{x} \, dx \ .$$

Le ipotesi del corollario 1.17 sono verificate in ogni rettangolo contenuto nel primo quadrante. Abbiamo quindi

$$\Phi'(y) = \frac{\exp(-y^{3/2})}{\sqrt{y}} \frac{1}{2\sqrt{y}} - \int_1^{\sqrt{y}} e^{-xy} dx =$$
$$= \frac{3 \exp(-y^{3/2})}{2y} - \frac{\exp(-y)}{y} \ .$$

Esercizi

1. Dimostrare la proposizione 1.7.
(Si assuma, per semplicità, che f abbia un'unica discontinuità in $c \in [a,b]$. Si consideri un intorno $[\alpha, \beta]$ di c abbastanza piccolo perché $(\beta - \alpha) \cdot$ (oscillazione di f in $[\alpha, \beta]$) $< \epsilon$ prefissato. Si trovi poi un'opportuna suddivisione di $[a,b]$ e si utilizzi il teorema 1.3).

2. Sia $f : [a,b] \to \mathbb{R}$ una funzione diversa da zero solo in un numero finito di punti. Dimostrare che f è integrabile e che il suo integrale è nullo.

3. Dimostare che, se si modifica, in un numero finito di punti, il valore di una funzione integrabile, si ottiene ancora una funzione integrabile con lo stesso valore dell'integrale.

4. Sia f continua in $[a,b]$, $f(x) \geq 0$, $\int_a^b f(x)dx = 0$. Allora segue che $f(x) = 0$ $\forall x \in [a,b]$.
(Per assurdo: se in un punto $\xi \in [a,b]$ fosse $f(\xi) > 0$, esisterebbe un intorno $[\xi - \delta, \xi + \delta]$ di ξ in cui sarebbe $f(x) > 0$...).

5.* Dimostrare che, se f e g sono integrabili su $[a,b]$, anche fg e f/g (se $|g(x)| \geq k > 0$ in $[a,b]$) sono integrabili.

6. *Teorema della media ponderata.* Siano $f, \varphi : [a,b] \to \mathbb{R}$ continue, $\varphi(x) \geq 0$ in $[a,b]$. Allora vale la formula

$$\int_a^b f(x)\varphi(x)dx = f(\xi)\int_a^b \varphi(x)dx$$

per un certo punto $\xi \in [a,b]$.

7. Data la funzione $f(t) = A\cos(\omega t + \varphi)$ periodica di periodo $T = \dfrac{2\pi}{\omega}$ (A, ω, φ parametri reali, $A, \omega > 0$), calcolare il suo valor medio su un intervallo $[0, nT + \tau]$ ($n \in \mathbb{N}, \tau \geq 0$); distinguere il caso $\tau = 0$ dal caso $\tau > 0$; mostrare che in ogni caso il valor medio tende a 0 quando $n \to +\infty$.

Calcolare la radice quadrata del valor medio della funzione $f(t)^2$ sullo stesso intervallo e mostrare che essa tende a $A/\sqrt{2}$ per $n \to +\infty$.

8.* Dimostrare che, se $f \in \mathcal{R}(a,b)$, ogni funzione integrale $F \in C^0([a,b])$.

9. *Formula di Taylor con resto in forma integrale.*

Sia $f \in C^{n+1}([a,b])$, $x, x_0 \in (a,b)$, $T_n(x)$ il polinomio di Taylor generato da f con centro in x_0, $E_n(x)$ il resto n–esimo. Posto

$$\Phi(t) = \sum_{k=0}^{n} \frac{(x-x_0)^k}{k!}(1-t)^k f^{(k)}(tx + (1-t)x_0)$$

abbiamo $\Phi(1) = f(x)$, $\Phi(0) = T_n(x)$, cosicché

$$E_n(x) = \Phi(1) - \Phi(0) = \int_0^1 \Phi'(t)dt\ .$$

Calcolando $\Phi'(t)$ (lo studente è invitato a sviluppare i calcoli) abbiamo

$$E_n(x) = \frac{(x-x_0)^{n+1}}{n!}\int_0^1 (1-t)^n f^{(n+1)}(tx + (1-t)x_0)dt\ . \tag{1.48}$$

10. Dimostrare, utilizzando il teorema di Liouville, che la funzione e^x/x non ammette primitiva elementare. Anche la funzione $1/\log x$ non è integrabile elementarmente (infatti, con una sostituzione, ci si riconduce al caso precedente); l'integrale

$$\int_0^x \frac{dt}{\log t} \qquad (0 < x < 1)$$

definisce una nuova funzione trascendente, detta *logaritmo integrale*.

11. Dimostrare che le funzioni $\dfrac{\sin x}{x}, \dfrac{\cos x}{x}, \sin x^2, \cos x^2$, non hanno primitive elementari.

(Ricordare la formula di Eulero: $e^{iy} = \cos y + i \sin y$; utilizzarla una volta con $y = x$, una volta con $y = x^2$).

12. *Irrazionalità di* π.

L'irrazionalità di π è stata dimostrata per la prima volta da J. H. Lambert nel 1768. La dimostrazione che segue è tratta dal libro "Irrational numbers" dell'americano Ivan Niven (1956). Al lettore chiediamo di completare i dettagli.

Consideriamo la funzione

$$f(x) = \frac{x^N(1-x)^N}{N!} ,$$

dove N è un intero positivo. Si può scrivere, utilizzando lo sviluppo di un binomio:

a) $f(x) = \frac{1}{N!} \sum_{n=N}^{2N} c_n x_n$ dove $c_n = (-1)^{n-N} \binom{n-N}{N}$.

Si ha:

b) se $N \le k \le 2N$, $f^{(k)}(x) = \frac{1}{N!} \sum_{n=k}^{2n} n(n-1)\dots(n-k+1)c_n x^{n-k}$ e dunque $f^{(k)}(0) = k! c_k / N!$; mentre se $0 \le k < N$, $f^{(k)}(0) = 0$.

c) $f^{(k)}(x) = (-1)^k f^{(k)}(1-x)$; dunque $f^{(k)}(1) = (-1)^k k! c_k / N!$ per $N \le k \le 2N$ mentre $f^{(k)}(1) = 0$ se $0 \le k \le N$.

Da *b*) e *c*) segue che $f^{(k)}(0)$ ed $f^{(k)}(1)$ sono numeri interi per ogni $k, 0 \le k \le 2N$. Ragioniamo ora per assurdo ponendo $\pi^2 = \frac{p}{q}$ con p e q numeri interi.

Costruiamo le due funzioni

$$F(x) = q^N \sum_{j=0}^{N} (-1)^j \pi^{2N-2j} f^{(2j)}(x) = \sum_{j=0}^{N} (-1)^j p^{N-j} q^j f^{(2j)}(x)$$

e

$$g(x) = F'(x) \sin \pi x - \pi F(x) \cos \pi x .$$

Queste funzioni hanno le seguenti proprietà:

d) $F(1)$ e $F(0)$ sono interi; $F''(x) = -\pi^2 [F(x) - p^N f(x)]$;

e) $g(1) = \pi F(1)$, $g(0) = -\pi F(0)$;

f) $g'(x) = F''(x) \sin \pi x + \pi F'(x) \cos \pi x - \pi F'(x) \cos \pi x + \pi^2 F(x) \sin \pi x =$
$= [F''(x) + \pi^2 F(x)] \sin \pi x = \pi^2 p^N f(x) \sin \pi x$.

Da *d*), *e*) e *f*) otteniamo, per il teorema della media:

$$\pi [F(1) - F(0)] = g(1) - g(0) = \int_0^1 g'(x) dx = \pi^2 p^N f(\xi) \sin \pi \xi$$

dove ξ è un punto opportuno tra 0 ed 1. Dunque si ricava, essendo

$$0 < \sin \pi\xi \le 1 \quad e \quad f(\xi) < \frac{1}{N!} \ :$$

g) $0 < F(1) - F(0) < \pi \dfrac{p^N}{N!}.$

Dalla *g)* e dalla *d)* si ottiene un assurdo se N è abbastanza grande, poiché

$$\lim_{N \to +\infty} \frac{p^N}{N!} = 0 \ .$$

13. *Alcune formule di ricorrenza.*

Per calcolare $\int (\sin x)^2 dx$ si può procedere, integrando per parti, nel modo seguente:

$$\int (\sin x)^2 dx = \int \sin x \cdot \sin x dx = -\sin x \cos x + \int (\cos x)^2 dx =$$

$$= -\sin x \cos x + \int [1 - (\sin x)^2] dx = -\sin x \cos x + x - \int (\sin x)^2 dx \ .$$

Si ricava così

$$\int (\sin x)^2 dx = \frac{1}{2}(x - \sin x \cos x) \ .$$

Posto $I_{m,n}(x) = \int (\sin x)^m (\cos x)^n dx$ dimostrare le formule seguenti (dove m, n sono interi ≥ 0, non entrambi nulli), sviluppando il procedimento sopra indicato:

$$I_{m,n}(x) = -\frac{(\cos x)^{n+1}(\sin x)^{m-1}}{n+m} + \frac{m-1}{n+m} I_{m-2,n}(x)$$

$$I_{m,n}(x) = \frac{(\sin x)^{m+1}(\cos x)^{n-1}}{n+m} + \frac{n-1}{n+m} I_{m,n-2}(x) \ .$$

14. Dalle formule precedenti, posto $I_m = \int_0^{\pi/2} (\sin x)^m dx$, ricavare:

$$I_m = \frac{m-1}{m} I_{m-2} \qquad (m = 2, 3, \ldots) \ .$$

Essendo $I_0 = \pi/2, I_1 = 1$ ricavare che, per ogni intero $p \ge 1$,

$$I_{2p} = \frac{1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2p-1)}{2 \cdot 4 \cdot 6 \cdot \ldots \cdot 2p} \ \frac{\pi}{2}$$

$$I_{2p+1} = \frac{2 \cdot 4 \cdot 6 \cdot \ldots \cdot 2p}{1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2p+1)} \ .$$

15. Posto $J_m(x) = \int \frac{dx}{(1+x^2)^m}$ $(m = 2, 3, \ldots)$, dimostrare la formula ricorrente

$$J_m(x) = \frac{3-2m}{2-2m} J_{m-1}(x) - \frac{x}{2(1-m)(1+x^2)^{m-1}}$$

(scrivere $\int \frac{dx}{(1+x^2)^m} = \int \frac{dx}{(1+x^2)^{m-1}} - \int \frac{x^2 dx}{(1+x^2)^m}$ e integrare per parti il secondo integrale).

16. Trovare formule di ricorrenza per gli integrali indefiniti

$$\int x^m \sin x dx \ , \ \int x^m \cos x dx \ , \ \int (\log x)^m dx \ , \ \int x^m e^x dx \ , \int x^n (\log x)^m dx \ .$$

17. Calcolare gli integrali:

$$\int \operatorname{tg} x dx \ , \ \int \operatorname{cotg} x dx \ , \ \int \frac{dx}{x \log x} \ , \ \int \frac{2ax+b}{ax^2+bx+c} dx$$

(tutti questi integrali si calcolano immediatamente osservando che:

$$\int \frac{\varphi'(x)}{\varphi(x)} dx = \int \frac{d}{dx} \log|\varphi(x)| dx = \log|\varphi(x)| \ ) \ .$$

18. *Integrali di funzioni razionali.*

Abbiamo visto in 4.3.2 che ogni funzione razionale può decomporsi nella somma di un polinomio e di frazioni semplici del tipo

$$\frac{A}{(x-a)^m} \qquad \frac{Mx+N}{(x^2+px+q)^m} \ .$$

L'integrazione delle funzioni del primo tipo è immediata:

$$\int \frac{Adx}{x-a} = A \log|x-a|$$

$$\int \frac{Adx}{(x-a)^m} = \frac{-A}{(m-1)(x-a)^{m-1}} \quad (m > 1) \ . \tag{1.49}$$

Le frazioni del secondo tipo possono scriversi nella forma

$$\frac{Mx+N}{[(x-\alpha)^2+\beta^2]^m}$$

avendo posto $p = -2\alpha$ e $q = \alpha^2 + \beta^2$ (si ricordi che $x^2 + px + q$ è irriducibile sul campo reale, perciò è $q > 0$). Con la sostituzione $x = \alpha + \beta t$, gli integrali di queste frazioni si riconducono ad integrali dei due tipi seguenti:

$$\int \frac{tdt}{(1+t^2)^m} \quad \int \frac{dt}{(1+t^2)^m} \ .$$

Il primo di questi due integrali, posto $r = t^2$, si riduce ancora ad uno del tipo (1.49); per il secondo abbiamo ricavato, nell'esercizio 15, la formula ricorrente; per $m = 1$ questo integrale è $\operatorname{arctg} t$.

Resta così dimostrata la proposizione 1.12.

Calcolare i seguenti integrali:

$$\int \frac{dx}{x^4-1} \qquad \left[\frac{1}{4}\log\left(\left|\frac{x-1}{x+1}\right|\right) - \frac{1}{2}\operatorname{arctg} x\right]$$

$$\int \frac{dx}{x^4+1} \qquad \left[\frac{\sqrt{2}}{4}(\operatorname{arctg}(\sqrt{2}x-1) + \operatorname{arctg}(\sqrt{2}x+1)) + \right.$$

$$\left. + \frac{1}{4\sqrt{2}}\log\left(\frac{x^2+\sqrt{2}x+1}{x^2-\sqrt{2}x+1}\right)\right]$$

$$\int \frac{dx}{(x-1)^2(x^2+1)} \qquad \left[\frac{1}{2(1-x)} - \frac{1}{2}\log|1-x| + \frac{1}{4}\log(1+x^2)\right].$$

Integrali di funzioni algebriche.

Si cerca, in generale, mediante opportune sostituzioni, di ricondurre l'integrale dato ad un altro con integranda razionale.

Daremo solo alcuni esempi; $R(x, y \dots)$ indicherà sempre una funzione razionale degli argomenti.

19. *Integranda del tipo* $R\left(x, \sqrt[r]{\frac{\alpha x+\beta}{\gamma x+\delta}}, \sqrt[s]{\frac{\alpha x+\beta}{\gamma x+\delta}}, \dots\right)$.

Questi integrali sono tutti razionalizzabili. Basta porre

$$\frac{\alpha x+\beta}{\gamma x+\delta} = t^k$$

essendo k il minimo comune multiplo degli indici di radice $r, s, \dots$

Calcolare i seguenti integrali

$$\int \frac{dx}{\sqrt{x}+\sqrt[3]{x}} = \begin{bmatrix} x = t^6 \\ dx = 6t^5 dt \end{bmatrix} = 6\int \frac{t^5 dt}{t^3+t^2} = \dots$$

$$\int \sqrt{\frac{x}{x-a}}\,dx = \begin{bmatrix} x/(x-a) = t^2 \\ x = at^2/(t^2-1) \end{bmatrix} = -2a\int \frac{t^2 dt}{(t^2-1)^2} = \dots$$

$$\int \frac{x+\sqrt{x-1}}{x-2\sqrt{x-1}}\,dx = \dots$$

20. *Integranda del tipo* $R(x, \sqrt{ax^2+bx+c})$.

Questi integrali sono abeliani, appartenenti alla conica di equazione: $y^2 - ax^2 - bx - c = 0$. Essi sono tutti razionalizzabili, mediante una opportuna sostituzione. Per evitare l'uso di espressioni immaginarie, conviene distinguere due casi.

i) Il trinomio ax^2+bx+c ha due radici reali e distinte λ, μ, cioè si può scrivere: $ax^2+bx+c=a(x-\lambda)(x-\mu)$. Si pone allora;

$$a\frac{x-\lambda}{x-\mu}=t^2, \text{ cioè } a(x-\lambda)(x-\mu)=t^2(x-\mu)^2 ,$$

da cui

$$\sqrt{ax^2+bx+c}=t(x-\mu) .$$

ii) Il trinomio non ha radici reali; lo supporremo sempre positivo (essendo sotto radice); ciò comporta che a e c siano positivi. Allora ognuna delle seguenti sostituzioni razionalizza l'integrale:

$$\sqrt{ax^2+bx+c}=t-\sqrt{a}x$$

$$\sqrt{ax^2+bx+c}=tx-\sqrt{c} .$$

Queste sostituzioni sono applicabili anche nel caso *i*) se uno dei due coefficienti, a o c, è positivo.

Il caso in cui il trinomio ha due radici reali coincidenti è banale, poiché, essendo: $ax^2+bx+c=a(x-\lambda)^2$, l'integranda è razionale.

Ricavare le formule seguenti:

$$\int\sqrt{x^2+a}\,dx=\frac{x}{2}\sqrt{x^2+a}+\frac{a}{2}\log\left(x+\sqrt{x^2+a}\right)$$

$$\int\frac{x^2dx}{\sqrt{x^2+a}}dx=\frac{x}{2}\sqrt{x^2+a}-\frac{a}{2}\log\left(x+\sqrt{x^2+a}\right)$$

$$\int\sqrt{a^2-x^2}\,dx=\frac{x}{2}\sqrt{a^2-x^2}+\frac{a^2}{2}\arcsin\frac{x}{a}$$

$$\int\frac{x^2dx}{\sqrt{a^2-x^2}}dx=-\frac{x}{2}\sqrt{a^2-x^2}+\frac{a^2}{2}\arcsin\frac{x}{a} .$$

21. Data l'iperbole di equazione: $\dfrac{x^2}{a^2}-\dfrac{y^2}{b^2}=1$ (vedi fig. 8.6) calcolare l'area $A(t)$ della parte grigia. Si trova:

$$A(t)=\frac{ab}{2}\left\{\frac{t}{a}\sqrt{\frac{t^2}{a^2}-1}-\log\left(\frac{t}{a}+\sqrt{\frac{t^2}{a^2}-1}\right)\right\}$$

Allora l'area $S(t)$ del settore OPV risulta:

$$S(t)=\frac{1}{2}ab\log\left(\frac{t}{a}+\sqrt{\frac{t^2}{a^2}-1}\right)=\frac{ab}{2}\,\mathrm{Ch}^{-1}\left(\frac{t}{a}\right) .$$

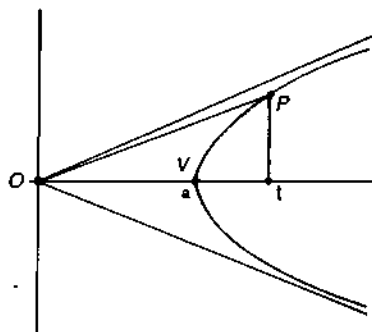



Fig. 8.6

Ciò giustifica il nome di: settore coseno iperbolico (e la notazione Sett Ch) dato alla funzione Ch^{-1} (e gli analoghi nomi dati alle altre funzioni iperboliche inverse: ved 5.3.5).

22. *Integrali di differenziali binomi.*

Si tratta di integrali del tipo: $\int x^m(a+bx^n)^p dx$, dove m, n, p sono numeri razionali. Cebicev (*) ha dimostrato che questi integrali sono calcolabili elementarmente se e solo se uno dei tre numeri

$$p\ ,\quad \frac{m+1}{n}\ ,\quad \frac{m+1}{n}+p$$

è intero. La sostituzione che rimuove la irrazionalità dell'integranda è, per ognuno dei tre casi sopra indicati, la seguente:

i) p intero: $x = t^k$, dove k è il minimo comune multiplo dei denominatori di m e n. Questo caso rientra in quelli trattati nell'esercizio 19.

ii) $\dfrac{m+1}{n}$ intero: $a + bx^n = t^{p_2}$, dove p_2 è il denominatore di $p\left(p = \dfrac{p_1}{p_2}\right)$.

iii) $\dfrac{m+1}{n} + p$ intero: $b + ax^n = t^{p_2}$.

Calcolare i seguenti integrali

$$\int \frac{\sqrt[3]{x}dx}{\sqrt[3]{(1+\sqrt[5]{x})^2}} \qquad m=\frac{1}{3}, n=\frac{1}{5}, p=-\frac{2}{3};\ \ \frac{m+1}{n}+p=6$$

$$\int x\sqrt[3]{1+x^3}dx \qquad m=1, n=3, p=\frac{1}{3};\ \ \frac{m+1}{n}+p=1$$

$$\int \frac{\sqrt[3]{1+\sqrt[4]{x}}}{\sqrt{x}}dx \qquad m=-\frac{1}{2}, n=\frac{1}{4}, p=\frac{1}{3};\ \ \frac{m+1}{n}=2\ .$$

Integrali di funzioni trascendenti.

23. Integrali con integranda del tipo $R(e^{ax})$ si razionalizzano con la sostituzione: $e^{ax}=t$. Calcolare gli integrali:

$$\int \frac{1+e^x}{e^{2x}+1}dx\ ,\quad \int \frac{\mathrm{Sh}\, x+1}{\mathrm{Sh}\, x-1}dx\ ,\quad \int e^{2x}\,\mathrm{Ch}\,\frac{x}{3}dx$$

24. *Integrali del tipo* $\int R(\sin x, \cos x)dx$ *con* $R(x, y)$ *funzione razionale.*

(*) P. Cebicev (1821–1894), matematico russo.

Per questi integrali la sostituzione: $x = 2\,\mathrm{arctg}\,t$ consente di ricondursi ad integrali di funzioni razionali. Infatti, abbiamo

$$t = \mathrm{tg}\,\frac{x}{2}\,,$$

$$\sin x = 2\sin\frac{x}{2}\cos\frac{x}{2} = 2\frac{\mathrm{tg}\,\frac{x}{2}}{1+(\mathrm{tg}\,\frac{x}{2})^2} = \frac{2t}{1+t^2}$$

$$\cos x = \left(\cos\frac{x}{2}\right)^2 - \left(\sin\frac{x}{2}\right)^2 = \frac{1-\left(\mathrm{tg}\,\frac{x}{2}\right)^2}{1+\left(\mathrm{tg}\,\frac{x}{2}\right)^2} = \frac{1-t^2}{1+t^2}$$

$$dx = \frac{2}{1+t^2}dt\,.$$

Perciò l'integrale dato si trasforma nell'integrale

$$2\int\frac{1}{1+t^2}R\left(\frac{2t}{1+t^2},\frac{1-t^2}{1+t^2}\right)dt$$

il quale ha una integranda razionale.

Verificare le formule

$$\int\frac{dx}{\sin x} = \log\left|\mathrm{tg}\,\frac{x}{2}\right|$$

$$\int\frac{dx}{\cos x} = \log\left|\mathrm{tg}\left(\frac{x}{2}+\frac{\pi}{4}\right)\right|$$

$$\int\frac{dx}{\sin x+\cos x} = \frac{1}{2}\left\{\log\left|\mathrm{tg}\left(\frac{x}{2}\right)-1+\sqrt{2}\right| - \log\left|\mathrm{tg}\left(\frac{x}{2}\right)-1-\sqrt{2}\right|\right\}.$$

25. L'integrale

$$\int\frac{dx}{a(\cos x)^2+b\sin x\cos x+c(\sin x)^2+d}$$

può razionalizzarsi con il metodo illustrato nell'esercizio 24; ma, osservando che la funzione integranda è periodica di periodo π, risulta più conveniente la seguente sostituzione: $x = \mathrm{arctg}\,t$, cioè

$$t = \mathrm{tg}\,x\ ,\ \ (\cos x)^2 = \frac{1}{1+t^2}\ ,\ \ \sin x\cos x = \frac{t}{1+t^2}\ ,\ \ (\sin x)^2 = \frac{t^2}{1+t^2}\,.$$

L'integrale dato diventa così

$$\int\frac{dt}{a+bt+ct^2+d(1+t^2)}$$

Possiamo così calcolare

$$\int \frac{dx}{(\cos x)^2} = \operatorname{tg} x$$

$$\int \frac{dx}{\sin x \cos x} = \log|\operatorname{tg} x|$$

$$\int \frac{dx}{(\sin x)^2} = -\operatorname{cotg} x\ .$$

26. Trovare per quali valori di m, n razionali l'integrale

$$\int (\sin x)^m (\cos x)^n dx$$

è calcolabile elementarmente.

(Fare la sostituzione $\sin x = t$ e ricordare il teorema di Cebicev sugli integrali dei differenziali binomi).

A conclusione di questa rassegna di metodi di razionalizzazione, vogliamo suggerire allo studente che, prima di applicare pedissequamente le sostituzioni consigliate nei vari casi sopra illustrati, osservi bene l'integrale proposto: spesso una scomposizione, una sostituzione bene azzeccata, una integrazione per parti, consentono di arrivare rapidamente al risultato.

I seguenti integrali, per esempio, vanno calcolati a mente in non più di $15''$:

$$\int \frac{e^x + \cos x}{e^x + \sin x} dx\ ,\quad \int \frac{2dx}{\sqrt{x} - \sqrt{x+2}}\ ,\quad \int \frac{\log\log x}{x} dx\ .$$

2. SERIE NUMERICHE

Prima di estendere l'operazione di integrazione alle funzioni non limitate e ad intervalli non limitati (sez. 3), conviene fermarci a studiare un'operazione più semplice che ha forti agganci con l'integrazione: l'operazione di somma estesa ad un numero infinito di addendi.

Già la definizione stessa di integrale come limite delle somme integrali al tendere a zero dell'ampiezza della suddivisione porta a considerare somme di questo tipo. Vediamo allora di precisare questo concetto e studiare le proprietà di questa operazione.

2.1 Definizione di serie e prime proprietà

Sia $\{a_n\}$, $n \in \mathbb{N}$, una successione di numeri complessi (in particolare, reali). Vogliamo dare significato all'operazione di somma degli infiniti addendi a_n.

A partire dalla successione $\{a_n\}$, costruiamo allora un'altra successione, $\{s_n\}$, i cui termini sono così definiti per ricorrenza:

$$s_0 = a_0 \quad , \quad s_{n+1} = s_n + a_{n+1} \tag{2.1}$$

e cioè

$$\begin{aligned} s_0 &= a_0 \\ s_1 &= a_0 + a_1 \\ &\vdots \\ s_n &= a_0 + a_1 + \ldots + a_n \\ &\vdots \end{aligned}$$

Definizione 2.1 - *Si dice* **serie** *di termini a_n la successione $\{s_n\}$ definita in* (2.1)*; l'elemento s_n viene detto somma parziale n–esima o ridotta n–esima della serie. La serie si dirà convergente, divergente, indeterminata (o irregolare) secondoché la successione $\{s_n\}$ sia convergente, divergente, indeterminata. Nel caso sia convergente, il limite di s_n, per $n \to +\infty$, si dirà* **somma** *della serie.*

La serie di termini a_n si indicherà con uno dei simboli seguenti:

$$a_0 + a_1 + \ldots + a_n + \ldots$$

oppure

$$\sum_{n=0}^{\infty} a_n$$

(o anche semplicemente $\sum a_n$, quando ciò non crei ambiguità).

Se la serie è convergente, ed A è la sua somma, cioè

$$\lim_{n \to +\infty} s_n = A \in \mathbf{C}$$

scriveremo

$$\sum_{n=0}^{\infty} a_n = A$$

anche se tale scrittura è un po' ambigua dal momento che, con lo stesso simbolo, si indica la serie e la sua somma.

Osserviamo che talvolta, invece di sommare a partire da 0, si parte da un indice $N > 0$; si scriverà allora $\sum\limits_{n=N}^{\infty} a_n$ e la successione delle ridotte sarà $\{s_n\}_{n \geq N}$. Osserviamo anche che l'indice di sommatoria, sopra indicato con n, è *muto*, come

la variabile di integrazione nell'integrale definito, e perciò potrà essere indicato con una qualsiasi altra lettera, anche nel corso di uno stesso calcolo.

Si usa talvolta scrivere, per un certo n,

$$\sum_{k=0}^{\infty} a_k = s_n + R_n .$$

La serie R_n, cioè $\sum_{k=n+1}^{\infty} a_k$, si dice *resto n–esimo* della serie $\sum_{k=0}^{\infty} a_k$; se questa è convergente, e A è la sua somma, il resto è un numero ben definito per ogni n

$$R_n = A - s_n$$

e risulta: $\lim_{n\to+\infty} R_n = 0$.

***Esempio* 2.1** - Sia $a_n = q^n$, $n \in \mathbb{N}$, $q \in \mathbb{C}$, la progressione geometrica (di ragione q) (cfr. esempi 4.3.5 e 4.4.14).

Abbiamo, se $q \neq 1$, utilizzando la formula (5.1) di 1.5.3 (valida anche nel caso di q complesso)

$$s_n = 1 + q + q^2 + \ldots + q^n = \frac{1 - q^{n+1}}{1 - q} .$$

Se invece $q = 1$ abbiamo

$$s_n = n .$$

Prendendo il limite, per $n \to +\infty$, otteniamo:

$$\lim_{n\to+\infty} s_n = \begin{cases} \dfrac{1}{1-q} & \text{se } |q| < 1 \\ \infty & \text{se } |q| > 1 \text{ oppure } q = 1 \\ \text{non esiste} & \text{se } |q| = 1 \quad (q \neq 1) . \end{cases}$$

Pertanto la serie (detta *serie geometrica* di ragione q)

$$\sum_{n=0}^{\infty} q^n \quad \text{è} \quad \begin{cases} \text{convergente con somma } \dfrac{1}{1-q} \text{ se } |q| < 1 \\ \text{divergente se } |q| > 1 \text{ oppure } q = 1 \\ \text{indeterminata se } |q| = 1 \ \ (q \neq 1) . \end{cases}$$

***Esempio* 2.2** - Sia $a_n = \dfrac{1}{n(n+1)}, n \geq 1$. Consideriamo la serie (detta *serie di Mengoli*):

$$\sum_{n=1}^{\infty} \frac{1}{n(n+1)} = \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \ldots + \frac{1}{n(n+1)} + \ldots$$

Poiché $\dfrac{1}{n(n+1)} = \dfrac{1}{n} - \dfrac{1}{n+1}$ la ridotta n–esima è

$$\begin{aligned} s_n &= \frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \ldots + \frac{1}{n(n+1)} = \\ &= \left(1 - \frac{1}{2}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \ldots + \left(\frac{1}{n} - \frac{1}{n+1}\right) = \\ &= 1 - \frac{1}{n+1} \end{aligned}$$

e quindi $s_n \to 1$ (per difetto) per $n \to +\infty$. Dunque la serie è convergente (per difetto) e ha per somma 1.

La serie considerata nell'esempio 2.2 rientra nella categoria delle cosiddette *serie telescopiche*, cioè serie del tipo:

$$\sum_{n=0}^{\infty} (a_n - a_{n+1}) \ .$$

Essendo, per queste serie,

$$s_n = \sum_{k=0}^{n} (a_k - a_{k+1}) = a_0 - a_{n+1}$$

risulta che il carattere della serie coincide con quello della successione $\{a_n\}$; se questa è convergente, la somma della serie è data da $a_0 - \lim_{n \to +\infty} a_n$.

Poiché l'operazione di somma di infiniti addendi è ben definita solo se la serie è convergente, è naturale cercare criteri che garantiscano l'esistenza del limite finito per la successione delle ridotte. Un criterio necessario e sufficiente è il criterio di Cauchy, dimostrato in 4.3.5 per le successioni a valori reali ed esteso in 4.4.3 anche alle successioni a valori complessi. Esso afferma che, per la convergenza della successione $\{s_n\}$, è necessario e sufficiente che, fissato $\varepsilon > 0$, esista un indice N (dipendente da ε) tale che risulti

$$|s_n - s_m| < \varepsilon \quad \text{se} \quad n, m \geq N \ .$$

Possiamo esprimere questa condizione direttamente sugli a_n; supponendo $n > m$ abbiamo infatti

$$|s_n - s_m| = |a_{m+1} + a_{m+2} + \ldots + a_n| .$$

Conviene allora porre: $m + 1 = p, n = p + q$ $(q \geq 0)$; otteniamo il seguente

▪ **Teorema 2.1** - (Criterio di Cauchy). *Condizione necessaria e sufficiente affinché la serie $\sum a_n$ sia convergente è che, fissato $\varepsilon > 0$, si possa determinare un indice N tale che, $\forall p \geq N$ e $\forall q \geq 0$ si abbia*

$$|a_p + a_{p+1} + \ldots + a_{p+q}| < \varepsilon . \tag{2.2}$$

***Esempio* 2.3** - Consideriamo la serie (detta *serie armonica*):

$$\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} + \ldots$$

Per un certo n, consideriamo le due somme parziali s_n e s_{2n}:

$$s_{2n} - s_n = \frac{1}{n+1} + \frac{1}{n+2} + \ldots + \frac{1}{2n} \quad (n \text{ addendi}) .$$

Poiché ogni addendo nell'espressione precedente è maggiore o uguale a $1/2n$, risulta:

$$s_{2n} - s_n \geq n\frac{1}{2n} = \frac{1}{2} .$$

Potendosi prendere n arbitrariamente grande, il criterio di Cauchy non è soddisfatto e perciò la serie armonica non converge (essa, come presto vedremo, diverge a $+\infty$).

Se consideriamo un caso particolare della condizione (2.2), ponendo in essa $q = 0$, otteniamo una *condizione solo necessaria* di convergenza, ma molto semplice e utile. Infatti la (2.2), per $q = 0$, diventa:

$$|a_p| < \varepsilon \qquad \forall\, p \geq N ,$$

condizione che possiamo esprimere anche nella forma del seguente

Corollario 2.2 - *Condizione necessaria affinché la serie $\sum a_n$ converga è che*

$$\lim_{n \to +\infty} a_n = 0 . \tag{2.3}$$

Per esempio, la condizione (2.3) consente subito di escludere che la serie geometrica converga per $|q| \geq 1$; infatti il termine generale q^n non tende a zero se $|q| \geq 1$. Che questa condizione non sia sufficiente per la convergenza è provato dal fatto che la serie armonica non converge!

Osservazione 2.1 - Prima di passare a considerare classi particolari di serie, ricordiamo che, dalle analoghe proprietà per le successioni, si deducono le seguenti per le serie:

i) il carattere di una serie (cioè il fatto di essere convergente, divergente, irregolare) non cambia se si alterano, si aggiungono, si tolgono un numero finito di addendi (ovviamente, se la serie è convergente, la sua somma viene alterata!);

ii) la convergenza della serie $\sum a_n$ equivale alla convergenza delle due serie (a termini reali) $\sum \text{Re}a_n$ e $\sum \text{Im}a_n$; se A_1 e A_2 sono le somme di queste due serie si ha

$$\sum_{n=0}^{\infty} a_n = A_1 + iA_2 \ .$$

D'ora in avanti considereremo quasi sempre serie a termini reali.

2.2 Serie a termini non negativi

Per le serie a termini (reali) non negativi si hanno alcuni semplici criteri sufficienti di convergenza. Cominciamo con l'osservare che la successione delle ridotte di una tale serie sarà crescente, poiché

$$s_{n+1} = s_n + a_{n+1} \geq s_n \ .$$

Perciò esiste $\lim\limits_{n \to +\infty} s_n$: tale limite è finito oppure $+\infty$ secondoché la successione $\{s_n\}$ è limitata oppure no. In ogni caso risulta: $\lim\limits_{n \to +\infty} s_n = \sup\limits_{n \in \mathbb{N}} \{s_n\}$.

Abbiamo perciò la seguente

Proposizione 2.3 - *Una serie $\sum a_n$ a termini non negativi o è convergente oppure è divergente a $+\infty$. Essa converge se e solo se la successione delle ridotte n–esime è limitata.*

Ricordando allora i teoremi di confronto per le successioni, abbiamo il seguente

▪ **Teorema 2.4** - (Criterio del confronto). *Siano $\sum a_n$ e $\sum b_n$ due serie a termini non negativi e tali che*

$$a_n \leq b_n \qquad \forall n \geq 0 \ . \tag{2.4}$$

Allora valgono le seguenti implicazioni

i) $\sum b_n$ convergente $\Rightarrow$ $\sum a_n$ convergente.

Dimostrazione - Sia $A_n = a_0 + a_1 + \ldots + a_n$ la ridotta n–esima della serie $\sum a_n$ e $B_n = b_0 + b_1 + \ldots + b_n$ la ridotta n–esima di $\sum b_n$. Per la (2.4) risulta $A_n \leq B_n$ $\forall n \in \mathbb{N}$.

Se allora $\sum b_n$ converge, la successione $\{B_n\}$ è limitata (proposizione 2.3); sia $B_n \leq B$ essendo $B = \sup\{B_n\}$ la somma di $\sum b_n$. Allora anche $\{A_n\}$ è limitata e perciò la serie $\sum a_n$ converge; detta $A = \sup\{A_n\}$ la sua somma, sarà $A \leq B$.

La *ii*) è equivalente alla *i*) essendone la proposizione contronominale. □

Esempi

2.4. La serie $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$, con α reale ≤ 1 è divergente.

Infatti l'affermazione è vera per $\alpha = 1$ (serie armonica); per $\alpha < 1$ la serie data è maggiorante della serie armonica, perciò diverge.

2.5. La serie $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$, con α reale ≥ 2 è convergente.

Infatti l'affermazione è vera per $\alpha = 2$ per confronto con la serie di Mengoli (esempio 2.2): $\dfrac{1}{n^2} \leq \dfrac{1}{n(n-1)}$; per $\alpha > 2$ si ha $\dfrac{1}{n^\alpha} \leq \dfrac{1}{n^2}$ e perciò la serie data è convergente.

Osservazione 2.2 - Per l'osservazione 2.1 *i*) possiamo applicare il criterio del confronto anche se la (2.4) non è verificata per tutti gli n, ma solo *definitivamente*, cioè da un certo valore di n in poi. Non è neanche necessario che tutti i termini della serie siano non negativi, essendo sufficiente che siano definitivamente non negativi. Questa osservazione ovviamente vale anche per gli altri criteri di questo paragrafo.

***Esempio* 2.6** - Consideriamo la serie

$$\sum_{n=0}^{\infty} \frac{n^2 - 10n}{n^4 + 1}$$

i cui termini sono positivi per $n \geq 10$. Poiché risulta

$$\frac{n^2 - 10n}{n^4 + 1} = \frac{1}{n^2} \cdot \frac{1 - 10/n}{1 + 1/n^4} \leq \frac{1}{n^2}$$

la serie data converge.

Il seguente corollario risulta utilissimo.

Corollario 2.5 - *Siano $\sum a_n$ e $\sum b_n$ due serie a termini positivi. Se $a_n \sim b_n$ allora le due serie hanno lo stesso carattere.*

Dimostrazione - È sufficiente osservare che, poiché $a_n/b_n \to 1$, definitivamente sarà

$$\frac{1}{2}b_n \le a_n \le \frac{3}{2}b_n \ . \ \square$$

***Esempio* 2.7** - Consideriamo le serie

$$a)\ \sum_{n=1}^{\infty} \sin\frac{1}{n},\quad b)\ \sum_{n=1}^{\infty} \log\left(1+\frac{1}{\sqrt{n}}\right),\quad c)\ \sum_{n=1}^{\infty} \frac{n^2+\log n+1}{n^4+n-1}\ .$$

Abbiamo

$a)$ $\sin\dfrac{1}{n} \sim \dfrac{1}{n} \Rightarrow$ la serie diverge

$b)$ $\log\left(1+\dfrac{1}{\sqrt{n}}\right) \sim \dfrac{1}{\sqrt{n}} \Rightarrow$ la serie diverge

$c)$ $\dfrac{n^2+\log n+1}{n^4+n-1} \sim \dfrac{1}{n^2} \Rightarrow$ la serie converge.

▪ **Teorema 2.6** - (Criterio della radice). *Sia $\sum a_n$ una serie a termini non negativi. Se esiste un numero l, $0 \le l < 1$, e un indice N per cui risulti*

$$\sqrt[n]{a_n} \le l \quad per \ \ n \ge N\ , \tag{2.5}$$

allora la serie è convergente. Se invece $\sqrt[n]{a_n} \ge 1$ per infiniti valori di n, la serie è divergente.

Dimostrazione - Se vale la (2.5), per $n \ge N$ si ha: $a_n \le l^n$ e la tesi segue confrontando la serie data con la serie geometrica di ragione l. Se invece $\sqrt[n]{a_n} \ge 1$, cioè $a_n \ge 1$ per infiniti valori di n, non è verificata la condizione necessaria di convergenza (2.3) e perciò la serie diverge. □

È utile il seguente

Corollario 2.7 - *Sia $\sum a_n$ una serie a termini non negativi. Se esiste il limite*

$$\lim_{n\to+\infty} \sqrt[n]{a_n} = l\ ,$$

allora, se $l > 1$ la serie diverge, se $l < 1$ la serie converge.

Dimostrazione - Sia $0 \le l < 1$. Per ogni $\varepsilon > 0$, esiste $N = N(\varepsilon)$ tale che, se $n \ge N$, si ha $\sqrt[n]{a_n} < l + \varepsilon$. Si rende allora $\varepsilon = (1-l)/2$; si avrà

$$\sqrt[n]{a_n} \le \frac{1+l}{2} < 1 \qquad \forall n \ge N$$

e perciò la serie converge.

Sia $l > 1$; allora, per infiniti n, deve essere $\sqrt[n]{a_n} > 1$ e perciò la serie diverge. □

Osservazione 2.3 - Se $l = 1$, ma *non per difetto*, allora ripetendo il ragionamento fatto sopra, ancora si conclude che la serie diverge. Se invece $\sqrt[n]{a_n} \to 1_-$, non si può, senza una ulteriore analisi, ricavare informazioni sulla convergenza della serie.

Se poi il limite non esiste, si può ricorrere al limite superiore. Sia

$$\bar{l} = \overline{\lim}\, \sqrt[n]{a_n} \, .$$

Con ragionamente identico a quello tenuto per l, si dimostra che, se $\bar{l} > 1$ la serie diverge, se $\bar{l} < 1$ la serie converge.

Esempi

2.8. Sia data $\sum_{n=1}^{\infty} \frac{a^n}{n^n}$ con $a \geq 0$, risulta, per $n \to +\infty$

$$\sqrt[n]{\frac{a^n}{n^n}} = \frac{a}{n} \to 0$$

e perciò la serie data converge.

2.9. Sia data $\sum_{n=1}^{\infty} n^{\alpha} a^n$, con $\alpha \in \mathbb{R}$, $a > 0$; risulta

$$\sqrt[n]{n^{\alpha} a^n} = a n^{\alpha/n} \, .$$

Calcoliamo $\lim_{n \to +\infty} n^{\alpha/n}$; basterà calcolare

$$\lim_{n \to +\infty} \log n^{\alpha/n} = \lim_{n \to +\infty} \frac{\alpha}{n} \log n = 0 \quad \forall \alpha \in \mathbb{R}$$

e perciò $\lim_{n \to +\infty} n^{\alpha/n} = 1$. Dunque, se $a < 1$ la serie converge. Se $a > 1$ la serie diverge. Se $a = 1$ la serie diventa $\sum_{n=1}^{\infty} n^{\alpha}$ e già sappiamo che per $\alpha \leq -2$ questa serie è convergente, per $\alpha \geq -1$ divergente.

▪ **Teorema 2.8** - (Criterio del rapporto). *Sia* $\sum a_n$ *una serie a termini positivi* ($a_n > 0$). *Se esiste un numero* l, $0 < l < 1$, *per cui risulti definitivamente*

$$\frac{a_{n+1}}{a_n} \leq l \tag{2.6}$$

allora la serie è convergente. Se definitivamente risulta $a_{n+1} \geq a_n$, *la serie diverge.*

Dimostrazione - Supponiamo pure, come è lecito, che la (2.6) valga per ogni $n \geq 0$; abbiamo

$$a_1 \leq l\, a_0$$
$$a_2 \leq l\, a_1 \leq l^2\, a_0$$
$$\vdots$$
$$a_n \leq l^n\, a_0\,.$$
$$\vdots$$

Confrontando allora la serie data con la serie geometrica di ragione l (si noti che le serie $\sum l^n$ e $\sum l^n a_0$ hanno ovviamente lo stesso carattere, essendo $a_0 \neq 0$) si ricava che la serie $\sum a_n$ è convergente.

Se invece risulta $a_{n+1} \geq a_n$ definitivamente la serie non può convergere (e perciò diverge) non essendo soddisfatto il criterio necessario di convergenza (2.3). □

Con dimostrazione analoga a quella fatta per il criterio della radice si può provare il seguente

Corollario 2.9 - *Sia $\sum a_n$ una serie a termini positivi. Se esiste il limite*

$$\lim_{n \to +\infty} \frac{a_{n+1}}{a_n} = l$$

allora, se $l > 1$ la serie diverge, se $l < 1$ la serie converge.

Osservazione 2.4 - Se $l = 1_+$, si può ancora concludere che la serie diverge, mentre negli altri casi il criterio non fornisce informazioni sul carattere della serie.

Se il limite non esiste, possiamo ancora considerare il limite superiore

$$\bar{l} = \overline{\lim_{n \to +\infty}}\, \frac{a_{n+1}}{a_n}\,.$$

Come prima, si dimostra che, se $\bar{l} < 1$, la serie converge. Ma se $\bar{l} \geq 1$ non si può, in generale, nulla concludere circa la divergenza della serie. Infatti sia, per esempio,

$$a_n = \begin{cases} \dfrac{1}{n^2} & \text{se } n \text{ è pari} \\[2mm] \dfrac{1}{n^3} & \text{se } n \text{ è dispari}\,. \end{cases}$$

La serie $\sum_{n=1}^{\infty} a_n$ è convergente, poiché minorante della serie $\sum_{n=1}^{\infty} \frac{1}{n^2}$, ma

$$\overline{\lim_{n \to +\infty}}\, \frac{a_{n+1}}{a_n} = +\infty\,.$$

***Esempio* 2.10** - La serie $\sum_{n=0}^{\infty} \frac{1}{n!}$ è convergente. Infatti:

$$\lim_{n\to+\infty} \frac{1/(n+1)!}{1/n!} = \lim_{n\to+\infty} \frac{1}{n+1} = 0 \, .$$

Possiamo dimostrare la convergenza della serie in un altro modo, che ci permette anche di calcolarne la somma. Scriviamo la formula di Mac Laurin per la funzione e^x con $x = 1$:

$$e = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \ldots + \frac{1}{n!} + \frac{e^\theta}{(n+1)!}$$

con $0 \leq \theta \leq 1$; la formula è valida $\forall n \geq 0$, essendo e^x infinitamente differenziabile. Abbiamo perciò, detta s_n la ridotta n–esima della serie data:

$$0 < e - s_n = \frac{e^\theta}{(n+1)!} < \frac{e}{(n+1)!} \, . \tag{2.7}$$

Prendendo il limite, per $n \to +\infty$, nella (2.7), si trova

$$e = \sum_{n=0}^{\infty} \frac{1}{n!} \, .$$

Si osservi anche che la serie converge ad e per difetto.

***Esempio* 2.11** - La serie $\sum_{n=1}^{\infty} n!/n^n$ è convergente. Infatti

$$\frac{(n+1)!}{(n+1)^{n+1}} \cdot \frac{n^n}{n!} = \left(\frac{n}{n+1}\right)^n = \left(1 - \frac{1}{n+1}\right)^n \to \frac{1}{e} < 1$$

per $n \to +\infty$.

Sappiamo che la serie $\sum 1/n^\alpha$ è convergente se $\alpha \geq 2$, divergente se $\alpha \leq 1$; la lacuna corrispondente ai valori di α compresi tra 1 e 2 può essere colmata ricorrendo al seguente criterio di convergenza, dovuto a Cauchy. (Si noti che, per la serie in esame, né il criterio della radice, né il criterio del rapporto, permettono di decidere circa la convergenza).

■ **Teorema 2.10** - (Criterio di condensazione). *Sia* $\{a_n\}$ *una successione a termini non negativi e decrescenti:*

$$a_0 \geq a_1 \geq \ldots \geq a_n \geq \ldots \geq 0 \, ;$$

allora $\sum a_n$ *converge se e solo se converge* $\sum 2^n a_{2^n}$.

Dimostrazione - Indichiamo con s_n la somma parziale della serie $\sum a_n$ e con σ_n la somma parziale della $\sum 2^n a_{2^n}$. Si ha

$$\begin{aligned}
&a_0 \leq a_0\\
&a_1 \leq a_0\\
&a_2 + a_3 \leq 2a_2\\
&a_4 + a_5 + a_6 + a_7 \leq 4a_4\\
&\dots\\
&a_{2^n} + a_{2^n+1} + \dots + a_{2^{n+1}-1} \leq 2^n a_{2^n}
\end{aligned}$$

e perciò, sommando,

$$s_{2^{n+1}-1} \leq a_0 + \sigma_n \ .$$

Analogamente abbiamo

$$\begin{aligned}
&a_0 \geq \frac{1}{2}a_0\\
&a_1 + a_2 \geq \frac{1}{2}2a_2\\
&a_3 + a_4 \geq \frac{1}{2}4a_4\\
&a_5 + a_6 + a_7 + a_8 \geq \frac{1}{2}8a_8\\
&\dots\\
&a_{2^{n-1}+1} + a_{2^{n-1}+2} + \dots + a_{2^n} \geq \frac{1}{2}2^n a_{2^n}
\end{aligned}$$

e perciò, sommando,

$$s_{2^n} \geq \frac{1}{2}\sigma_n \ .$$

Si vede così che, se la successione $\{\sigma_n\}$ è limitata, lo stesso vale per la $\{s_n\}$ e viceversa; perciò le due serie corrispondenti sono o entrambe convergenti, o entrambe divergenti. □

***Esempio* 2.12** - La serie $\displaystyle\sum_{n=1}^{\infty} \frac{1}{n^\alpha}$ è convergente per $\alpha > 1$, divergente (come già sappiamo) per $\alpha \leq 1$.

Poiché la successione $1/n^\alpha$ è decrescente (per ogni $\alpha > 0$) applichiamo il teorema precedente. Abbiamo

$$2^n a_{2^n} = 2^n \frac{1}{2^{\alpha n}} = (2^{1-\alpha})^n \ .$$

Ma la serie geometrica di ragione $2^{1-\alpha}$ converge se $\alpha > 1$ e diverge se $\alpha \leq 1$; lo stesso allora accadrà alla serie di partenza.

2.3 Convergenza e convergenza assoluta

Torniamo allo studio delle serie a termini di segno qualsiasi, o addirittura a termini complessi. I criteri visti in precedenza per le serie a termini positivi possono applicarsi anche a serie con termini qualsiasi, grazie al seguente teorema.

Data la serie $\sum a_n$, consideriamo la serie dei moduli, $\sum |a_n|$; questa è una serie a termini reali non negativi.

■ **Teorema 2.11** - *Se la serie $\sum |a_n|$ è convergente, allora anche la serie $\sum a_n$ è convergente.*

Dimostrazione - Per la disuguaglianza triangolare abbiamo, $\forall p, q \geq 0$

$$|a_p + a_{p+1} + \ldots + a_{p+q}| \leq |a_p| + |a_{p+1}| + \ldots + |a_{p+q}| \, .$$

La convergenza di $\sum |a_n|$ implica che, fissato $\varepsilon > 0$, esista N per cui la somma a destra della formula precedente sia $< \varepsilon$, $\forall p \geq N, \forall q \geq 0$; ma allora ciò è vero anche per il termine a sinistra; questo implica, per il criterio di Cauchy, la convergenza della serie $\sum a_n$. □

Come presto vedremo, il teorema non è invertibile; cioè può accadere che la serie $\sum a_n$ sia convergente, ma non la serie $\sum |a_n|$. Conviene allora indicare con un nome apposito la circostanza che sia l'una che l'altra serie siano convergenti.

Definizione 2.2 - *Una serie $\sum a_n$ si dirà assolutamente convergente se converge la serie $\sum |a_n|$.*

Dunque la convergenza assoluta implica la convergenza (ordinaria); il viceversa non è vero.

Esempi

2.13. La serie $\sum_{n=1}^{\infty} (-1)^n/n^\alpha$, con $\alpha > 1$ è convergente, anzi assolutamente convergente. Infatti si ha

$$\left| \frac{(-1)^n}{n^\alpha} \right| = \frac{1}{n^\alpha}$$

e perciò l'affermazione discende dall'esempio 2.12. Se $\alpha = 1$ nulla, per il momento, possiamo concludere circa la convergenza della serie data, poiché la serie dei valori assoluti corrispondente è la serie armonica, che è divergente. Nell'esempio 2.15 mostreremo che $\sum_{n=1}^{\infty} (-1)^n/n$ è convergente, fornendo così un esempio di serie convergente, ma non assolutamente convergente.

2.14. Sia $z \in \mathbb{C}$. La serie

$$\sum_{n=0}^{\infty} \frac{z^n}{n!}$$

è detta *serie esponenziale*. Essa è assolutamente convergente $\forall z$. Infatti abbiamo, applicando il criterio del rapporto alla serie dei moduli,

$$\frac{|z|^{n+1}}{(n+1)!} \cdot \frac{n!}{|z|^n} = \frac{|z|}{n+1} \to 0 \text{ per } n \to +\infty, \ \forall z \in \mathbb{C}.$$

Mostreremo più avanti che la somma di questa serie è il numero complesso e^z; questo risultato generalizza quello dell'esempio 2.10.

Per le serie a termini complessi, o anche reali ma di segno qualsiasi, il criterio espresso dal teorema 2.11 è praticamente l'unico a disposizione per studiarne la convergenza (ma vedi anche il teorema 2.13).

Un cenno particolare meritano le *serie a termini di segno alternato*, cioè serie che si presentano sotto la forma

$$\sum_{n=0}^{\infty} (-1)^n a_n$$

dove gli a_n sono numeri reali positivi. Per queste serie vale il seguente criterio di convergenza.

■ **Teorema 2.12** - (Criterio di Leibniz). *Sia data la serie* $\sum_{n=0}^{\infty} (-1)^n a_n$, *con* $a_n > 0 \ \forall n \in \mathbb{N}$. *Se*

i) la successione $\{a_n\}$ *è decrescente*
ii) $\lim_{n \to +\infty} a_n = 0$

allora la serie è convergente. Inoltre, le somme parziali di indice pari approssimano la somma per eccesso, quelle di indice dispari per difetto; il resto della serie è maggiorato, in valore assoluto, dal primo termine trascurato.

Dimostrazione - Posto, come al solito,

$$s_n = a_0 - a_1 + a_2 - \ldots (-1)^n a_n$$

si osserva che, per l'ipotesi *i*) del teorema,

$$s_{2n+2} = s_{2n} - (a_{2n+1} - a_{2n+2}) \leq s_{2n}$$

$$s_{2n+1} = s_{2n-1} + (a_{2n} - a_{2n+1}) \geq s_{2n-1}$$

cosicché la successione delle ridotte di indice pari è decrescente, quella delle ridotte di indice dispari è crescente. Inoltre

$$s_{2n} - s_{2n+1} = a_{2n+1} \tag{2.8}$$

e perciò $s_{2n} \geq s_{2n+1} \geq s_{2n-1} \geq \ldots \geq s_1$: la successione $\{s_{2n}\}$ è limitata inferiormente; perciò converge: sia S il suo limite (che coincide con il suo estremo inferiore). Dalla (2.8), facendo tendere n all'infinito e tenendo presente l'ipotesi *ii*), ricaviamo che anche la successione $\{s_{2n+1}\}$ tende allo stesso limite S; perciò la serie converge ed S è la sua somma. È evidente inoltre, per quanto detto, che $s_{2n} \to S$ per eccesso, $s_{2n+1} \to S$ per difetto. Valgono perciò le disuguaglianze

$$s_{2n-1} \leq S \leq s_{2n} \qquad s_{2n+1} \leq S \leq s_{2n}$$

dalle quali si deducono le seguenti:

$$0 \leq S - s_{2n-1} \leq s_{2n} - s_{2n-1} = a_{2n}$$

$$0 \leq s_{2n} - S \leq s_{2n} - s_{2n+1} = a_{2n+1}\,.$$

La figura 8.7 illustra il modo di convergere a S delle somme parziali. □

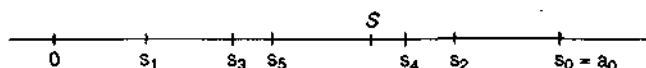



Fig. 8.7

***Esempio* 2.15** - Le serie

$$\sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \qquad \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}$$

sono entrambi convergenti, in virtù del teorema 2.12 (la prima è assolutamente convergente (esempio 2.10), la seconda no). La prima serie è rapidamente convergente; la sua somma è $1/e$, come facilmente si ottiene scrivendo la formula di Mac Laurin per la funzione e^x con $x = -1$ e procedendo come nell'esempio 2.10 (vedi anche l'esercizio 4). Sommando i primi 6 termini della serie

$$1 - 1 + \frac{1}{2} - \frac{1}{6} + \frac{1}{24} - \frac{1}{120} = \frac{11}{30} = 0,3\overline{6}$$

si ottiene un valore per difetto di $1/e$ con un errore che non supera $1/7! = 1/720$.

La seconda serie è lentamente convergente; per avere un valore della somma approssimato (per eccesso) a meno di $1/100$ bisogna sommare 99 termini! Come vedremo, la somma di questa serie è $\log 2$; ma la serie non è adatta al calcolo approssimato di questo numero.

Prima di concludere il paragrafo vediamo ancora un criterio di convergenza per serie a termini di segno qualsiasi (o addirittura complessi).

Siano $\{a_n\}$ e $\{b_n\}$ ($n \in \mathbb{N}$) due successioni, a valori reali o complessi; detta A_n la somma parziale n–esima relativa alla successione $\{a_n\}$, e posto, per comodità di scrittura, $A_{-1} := 0$, abbiamo, per ogni coppia di interi p, q con $0 \leq p \leq q$:

$$\sum_{n=p}^{q} a_n b_n = -\sum_{n=p}^{q-1} A_n(b_{n+1} - b_n) + A_q b_q - A_{p-1} b_p \,. \tag{2.9}$$

La (2.9) si ricava immediatamente scrivendo:

$$\sum_{n=p}^{q} a_n b_n = \sum_{n=p}^{q} (A_n - A_{n-1}) b_n = \sum_{n=p}^{q} A_n b_n - \sum_{n=p}^{q} A_{n-1} b_n = \ (^*)$$

$$\sum_{n=p}^{q} A_n b_n - \sum_{n=p-1}^{q-1} A_n b_{n+1} = \sum_{n=p}^{q-1} A_n (b_n - b_{n+1}) + A_q b_q - A_{p-1} b_p \,.$$

La formula (2.9) viene detta *formula di sommazione per parti*; l'analogia con la formula di integrazione per parti è evidente.

Dalla (2.9) si ricava il seguente

■ **Teorema 2.13** - *Siano $\{a_n\}$ e $\{b_n\}$ due successioni tali che:*

i) $\{a_n\}$ è a valori complessi ed esiste un numero positivo M tale che

$$|A_n| \leq M \qquad \forall n \geq 0$$

essendo A_n la ridotta n–esima di $\{a_n\}$

ii) $\{b_n\}$ è a valori reali positivi e tende monotonamente a zero

$$b_0 \geq b_1 \geq b_2 \geq \ldots \geq 0 \,, \quad \lim_{n \to +\infty} b_n = 0 \,.$$

Allora la serie $\sum_{n=0}^{\infty} a_n b_n$ è convergente.

Si osservi che, se $a_n = (-1)^n$ e i b_n sono reali, otteniamo come caso particolare il criterio di Leibniz per le serie a termini di segno alternato.

(*) Posto $m = n - 1$ abbiamo $\sum_{n=p}^{q} A_{n-1} b_n = \sum_{m=p-1}^{q-1} A_m b_{m+1}$; possiamo poi, come abbiamo osservato, cambiare il nome dell'indice di sommatoria, e scrivere ancora n in luogo di m.

Dimostrazione - Fissato $\varepsilon > 0$, sia n_0 tale che $b_p < \varepsilon$ se $p \geq n_0$. Presi allora p, q interi $\geq n_0$, $q \geq p$, dalla (2.9) si ha:

$$\begin{aligned}\left|\sum_{n=p}^{q} a_n b_n\right| &\leq \sum_{n=p}^{q-1} |A_n||b_{n+1} - b_n| + |A_q||b_q| + |A_{p-1}||b_p| \leq \\ &\leq M\left(\sum_{n=p}^{q-1}(b_n - b_{n+1}) + b_q + b_p\right) = \\ &= M(b_p - b_{p+1} + b_{p+1} - b_{p+2} + \ldots + b_{q-1} - b_q + b_q + b_p) = \\ &= 2b_pM < 2\varepsilon M \ .\end{aligned}$$

Pertanto la serie converge, per l'arbitrarietà di ε, in virtù del criterio di Cauchy. □

***Esempio* 2.16** - Consideriamo le serie: $\displaystyle\sum_{n=1}^{\infty} \frac{\sin n}{n}$ e $\displaystyle\sum_{n=1}^{\infty} \frac{\cos n}{n}$; per la formula di Eulero: $e^{in} = \cos n + i \sin n$ le due serie date sono, rispettivamente, parte immaginaria e parte reale della serie $\displaystyle\sum_{n=1}^{\infty} \frac{e^{in}}{n}$; applichiamo il criterio sopra illlustrato prendendo $a_n = e^{in}$ e $b_n = \dfrac{1}{n}$. Risulta allora

$$|A_n| = \left|\sum_{k=0}^{n} (e^i)^k\right| = \left|\frac{1 - e^{i(n+1)}}{1 - e^i}\right| \leq \frac{2}{|1 - e^i|} \ .$$

Pertanto la serie $\displaystyle\sum_{n=1}^{\infty} \frac{e^{in}}{n}$ converge; lo stesso accade della sua parte reale e della parte immaginaria, cioè delle serie sopra considerate. Si osservi anche che, essendo $\left|\dfrac{e^{in}}{n}\right| = \dfrac{1}{n}$, le serie date non convergono assolutamente.

2.4 Operazioni sulle serie

Siano $\sum a_n$ e $\sum b_n$ due serie (a termini reali qualsiansi) e sia $c \in \mathbb{R}$. Definiamo:

i) Prodotto di una serie per un numero

$$c \sum_{n=0}^{\infty} a_n := \sum_{n=0}^{\infty} c a_n$$

ii) Somma di due serie

$$\sum_{n=0}^{\infty} a_n + \sum_{n=0}^{\infty} b_n := \sum_{n=0}^{\infty} (a_n + b_n) .$$

Le affermazioni riguardanti il carattere delle serie così definite discendono immediatamente dai teoremi dimostrati nel calcolo dei limiti.

Proposizione 2.14 - *Se $c \neq 0$, la serie $\sum c a_n$ ha lo stesso carattere della serie $\sum a_n$; in particolare, se $\sum a_n$ converge e ha per somma A, anche $\sum c a_n$ converge e ha per somma cA.*

Se $c = 0$ la serie $\sum c a_n$ è convergente con somma 0 qualunque sia la serie $\sum a_n$.
In particolare (prendendo $c = -1$) si osservi che i criteri di convergenza visti per le serie a termini definitivamente non negativi possono applicarsi anche alle serie a termini definitivamente non positivi.

Proposizione 2.15 - *Se $\sum a_n$ e $\sum b_n$ sono convergenti (con somme A e B rispettivamente) anche $\sum (a_n + b_n)$ è convergente, con somma $A + B$.*

Lo studente non avrà difficoltà ad enunciare, quando ciò è possibile, proposizioni analoghe nel caso che una o entrambe le serie addende siano divergenti o indeterminate.

iii) Prodotto (secondo Cauchy) di due serie

$$\left(\sum_{n=0}^{\infty} a_n\right) \cdot \left(\sum_{n=0}^{\infty} b_n\right) := \sum_{n=0}^{\infty} c_n$$

dove si è posto

$$c_n = \sum_{k=0}^{n} a_k b_{n-k}$$

cioè, più esplicitamente,

$$\begin{aligned} c_0 &= a_0 b_0 \\ c_1 &= a_0 b_1 + a_1 b_0 \\ c_2 &= a_0 b_2 + a_1 b_1 + a_2 b_0 \\ &\vdots \\ c_n &= a_0 b_n + a_1 b_{n-1} + \ldots + a_{n-1} b_1 + a_n b_0 \end{aligned}$$

La definizione si giustifica osservando che essa ripete, estendendola, la usuale

regola di moltiplicazione per i polinomi:

$$(a_0 + a_1x + a_2x^2 + \ldots + a_nx^n)(b_0 + b_1x + b_2x^2 + \ldots + b_nx^n) =$$
$$= a_0b_0 + (a_0b_1 + a_1b_0)x + (a_0b_2 + a_1b_1 + a_2b_0)x^2 + \ldots +$$
$$+(a_0b_n + \ldots + a_nb_0)x^n \ .$$

■ **Teorema 2.16** - (di Mertens). *Siano $\sum a_n$ e $\sum b_n$ due serie convergenti, con somme A e B rispettivamente; sia inoltre una delle due, diciamo $\sum a_n$, assolutamente convergente. Allora la serie prodotto $\sum c_n$ è convergente con somma $C = AB$.*

Dimostrazione - Siano A_n, B_n, C_n le ridotte n–esime delle serie $\sum a_n, \sum b_n, \sum c_n$ rispettivamente. Osserviamo che possiamo scrivere

$$\begin{aligned} C_n &= a_0b_0 + (a_0b_1 + a_1b_0) + \ldots + (a_0b_n + \ldots + a_nb_0) = \\ &= a_0B_n + a_1B_{n-1} + \ldots + a_{n-1}B_1 + a_nB_0 = \\ &= a_0(B_n - B + B) + a_1(B_{n-1} - B + B) + \ldots + a_{n-1}(B_1 - B + B) + \\ &\quad + a_n(B_0 - B + B) = \\ &= A_nB - a_0R_n - a_1R_{n-1} - \ldots - a_{n-1}R_1 - a_nR_0 \end{aligned}$$

avendo indicato con $R_n = B - B_n$ il resto n–esimo della serie $\sum b_n$.

Mostriamo ora che

$$\lim_{n\to+\infty} \sum_{k=0}^{n} a_kR_{n-k} = 0 \ .$$

Poiché $A_nB \to AB$ per $n \to +\infty$ il teorema sarà dimostrato.

Fissiamo $\varepsilon > 0$ e determiniamo anzitutto $K = K(\varepsilon)$ tale per cui $|R_k| < \varepsilon$ se $k \geq K$; ciò è possibile, poiché $R_k \to 0$ per $k \to +\infty$.

Sia invece $L = \max\limits_{k\leq K}\{|R_k|\}$. Allora, per $n > K$ abbiamo:

$$\begin{aligned} \left|\sum_{k=0}^{n} a_kR_{n-k}\right| &= \left|\sum_{k=0}^{n-K} a_kR_{n-k} + \sum_{k=n-K+1}^{n} a_kR_{n-k}\right| \leq \\ &\leq \sum_{k=0}^{n-K} |a_k||R_{n-k}| + \sum_{k=n-K+1}^{n} |a_k||R_{n-k}| \leq \\ &\leq \varepsilon \sum_{k=0}^{n-K} |a_k| + L \sum_{k=n-K+1}^{n} |a_k| \ . \end{aligned}$$

Poiché la serie $\sum |a_k|$ converge (sia A^* la sua somma) sarà $\displaystyle\sum_{k=0}^{n-K} |a_k| \leq A^*$ e inoltre

(per il criterio di Cauchy) possiamo trovare un indice $N = N(\varepsilon)$ per cui, se $n \geq N$, risulti

$$\sum_{k=n-K+1}^{n} |a_k| < \varepsilon .$$

Avremo allora, per $n \geq N$,

$$\left|\sum_{k=0}^{n} a_k R_{n-k}\right| \leq \varepsilon A^* + \varepsilon L$$

che, per l'arbitrarietà di ε, prova la tesi. □

L'ipotesi di assoluta convergenza di una (almeno) delle due serie fattori non è rimuovibile (cfr. esercizio 10).

È facile poi mostrare che, se entrambi i fattori convergono assolutamente, anche la serie prodotto converge assolutamente.

2.5 Proprietà associativa e commutativa

Poiché le somme finite godono delle fondamentali proprietà associativa e commutativa, è naturale chiedersi se tali proprietà siano mantenute nel passaggio alle somme infinite.

Esaminiamo dapprima la proprietà associativa.

Sia data la serie

$$\sum_{n=0}^{\infty} a_n = a_0 + a_1 + a_2 + \ldots + a_n + \ldots$$

e raggruppiamo, con una legge qualsiasi, un certo numero di termini consecutivi; per esempio

$$a_0 + (a_1 + a_2 + a_3) + (a_4 + a_5) + a_6 + a_7 + (a_8 + a_9) + \ldots$$

Indichiamo poi con b_n i nuovi termini così ottenuti (nell'esempio sarebbe:

$$\begin{aligned} b_0 &= a_0 \\ b_1 &= a_1 + a_2 + a_3 \\ b_2 &= a_4 + a_5 \\ b_3 &= a_6 \\ b_4 &= a_7 \\ b_5 &= a_8 + a_9, \ldots) \end{aligned}$$

e consideriamo la serie $\sum_{n=0}^{\infty} b_n$. Se indichiamo con A_n e B_n rispettivamente le somme parziali n–esime di $\sum a_n$ e $\sum b_n$, risulta chiaro che $\{B_n\}$ è una sottosuccessione di $\{A_n\}$. Perciò, se $\{A_n\}$ converge, anche $\{B_n\}$ converge (allo stesso

limite); se $\{A_n\}$ diverge, anche $\{B_n\}$ diverge. Ciò porta a concludere che: *per le serie convergenti oppure divergenti vale la proprietà associativa.* Questa invece non vale per le serie indeterminate; sia infatti, per esempio,

$$\sum_{n=0}^{\infty}(-1)^n = 1 - 1 + 1 - 1 + \dots ; \tag{2.10}$$

la successione delle ridotte è: $1, 0, 1, 0, 1, 0, \dots$; perciò la serie è indeterminata. Se associamo i termini della serie (2.10) a coppie (i primi due, poi gli altri due, e così via) tutti i termini della nuova serie così ottenuta sono nulli e perciò questa converge con somma zero.

Passiamo ora alla proprietà commutativa. Definiamo anzitutto che cosa si intende per riordinamento dei termini di una serie.

Definizione 2.3 - *Diremo che una serie* $\sum_{n=0}^{\infty} b_n$ *è un riordinamento della serie* $\sum_{n=0}^{\infty} a_n$ *se esiste una applicazione* $j : \mathbb{N} \to \mathbb{N}$ *biunivoca tale che:*

$$b_n = a_{j(n)} .$$

Essendo j biunivoca, è chiaro che, se $\sum b_n$ è un riordinamento di $\sum a_n$, allora $\sum a_n$ è un riordinamento di $\sum b_n$.

Ci proponiamo di stabilire se il riordinamento di una serie ha lo stesso carattere (la stessa somma) della serie di partenza. Considereremo serie a termini reali.

■ **Teorema 2.17** - *Sia* $\sum a_n$ *una serie assolutamente convergente. Allora ogni suo riordinamento è pure assolutamente convergente e ha la stessa somma.*

Dimostrazione - Se $a_n \geq 0$, $\forall n \in \mathbb{N}$ (e quindi anche $b_n \geq 0\ \forall n \in \mathbb{N}$) posto $A = \sum a_n$, abbiamo (indicando, come al solito, con A_n e B_n le ridotte n–esime delle serie $\sum a_n$ e $\sum b_n$ rispettivamente)

$$B_n = b_0 + b_1 + \dots + b_n = a_{j(0)} + a_{j(1)} + \dots + a_{j(n)} \leq A ;$$

perciò la serie $\sum b_n$ è convergente e la sua somma B risulta $\leq A$. Ma scambiando i ruoli di $\sum a_n$ e $\sum b_n$, deduciamo che $A \leq B$, da cui segue $A = B$.

Sia ora a_n di segno qualsiasi. Poniamo

$$a_n^+ = \max\{a_n, 0\} \qquad b_n^+ = \max\{b_n, 0\}$$
$$a_n^- = \max\{-a_n, 0\} \quad b_n^- = \max\{-b_n, 0\}$$

per cui si ha

$$a_n = a_n^+ - a_n^- \qquad b_n = b_n^+ - b_n^-$$
$$|a_n| = a_n^+ + a_n^- \quad |b_n| = b_n^+ + b_n^-$$

e ancora

$$0 \leq a_n^+ \leq |a_n| \qquad 0 \leq b_n^+ \leq |b_n|$$
$$0 \leq a_n^- \leq |a_n| \qquad 0 \leq b_n^- \leq |b_n| \quad .$$

Si osservi che tutte le serie $\sum a_n^+, \sum a_n^-, \sum b_n^+, \sum b_n^-$ sono a termini non negativi, e che le serie $\sum b_n^+$ e $\sum b_n^-$ sono riordinamenti delle serie $\sum a_n^+$ e $\sum a_n^-$ rispettivamente.

La convergenza della serie $\sum |a_n|$ implica allora, per il teorema del confronto, la convergenza delle serie $\sum a_n^+$ e $\sum a_n^-$, e questo implica, per la prima parte della dimostrazione, la convergenza di $\sum b_n^+$ e $\sum b_n^-$ e perciò di $\sum |b_n|$. Essendo poi, per la prima parte della dimostrazione, $\sum a_n^+ = \sum b_n^+$ e $\sum a_n^- = \sum b_n^-$ segue che $\sum a_n = \sum b_n$. □

L'ipotesi di assoluta convergenza nel teorema 2.16 non può essere rimossa. Infatti vale il seguente

■ **Teorema 2.18** - (di Riemann–Dini) *Sia $\sum a_n$ una serie convergente, ma non assolutamente convergente. Allora, scelto comunque un numero $S \in \mathbb{R}$, esiste un riordinamento della serie data convergente con somma S; esistono anche riordinamenti di $\sum a_n$ che sono divergenti, altri che sono indeterminati.*

Dimostrazione - Consideriamo le serie $\sum a_n^+$ e $\sum a_n^-$ (utilizziamo le stesse notazioni del teorema precedente). Esse non possono essere entrambe convergenti (altrimenti lo sarebbe anche $\sum |a_n|$) né può essere una convergente e l'altra divergente (perché altrimenti sarebbe $\sum a_n$ divergente); dunque entrambe sono divergenti (a $+\infty$).

Siano ora $p_0, p_1, \dots p_n, \dots$ i termini non negativi della serie $\sum a_n$, elencati nell'ordine in cui compaiono, e $q_0, q_1, \dots, q_n \dots$ quelli negativi. Le serie $\sum p_n$ e $-\sum q_n$ differiscono dalle serie $\sum a_n^+$ e $\sum a_n^-$ per la presenza, in queste ultime, di altri termini nulli. Sarà dunque

$$\sum_{n=0}^{\infty} p_n = +\infty \quad , \quad \sum_{n=0}^{\infty} q_n = -\infty \ . \tag{2.11}$$

e inoltre, per la convergenza della serie $\sum a_n$,

$$\lim_{n \to +\infty} p_n = 0 \quad , \quad \lim_{n \to +\infty} q_n = 0 \ . \tag{2.12}$$

Sia dato $S \in \mathbb{R}$ e costruiamo un riordinamento $\sum b_n$ di $\sum a_n$ con le seguenti regole.

Poniamo $b_0 = p_0$. Poniamo $b_1 = q_0$ se $b_0 > S$, altrimenti poniamo $b_1 = p_1$. In generale, determinati i termini $b_0, b_1, \dots, b_n$, costituiti, diciamo, dai termini $p_0, p_1, \dots p_{k-1}$ e $q_0, q_1, \dots q_{n-k}$, sia $s_n = b_0 + b_1 + \dots + b_n$. Porremo allora $b_{n+1} = p_k$ se $s_n \leq S$, altrimenti $b_{n+1} = q_{n-k+1}$. Procedendo in tal modo non potrà mai accadere, a causa delle (2.11) che, da un certo posto in poi, risulti sempre $s_n > S$ oppure $s_n \leq S$, cosicché tutti i termini p_k e q_k, uno dopo l'altro, saranno scelti per costruire la serie $\sum b_n$ che risulterà così un riordinamento della $\sum a_n$. Inoltre, almeno da un certo posto in poi, la differenza $s_n - S$ sarà $\leq$ dell'ultimo dei termini non negativi $\{p_k\}$ considerati e $\geq$ dell'ultimo dei termini negativi $\{q_k\}$ considerati; in conseguenza delle (2.12) avremo allora

$$\lim_{n \to +\infty} s_n = S \ .$$

Con procedimento analogo, che non sviluppiamo in dettaglio, si possono costruire riordinamenti divergenti a $+\infty$ o a $-\infty$ o indeterminati. □

*2.6 Medie aritmetiche

In molte applicazioni della matematica, per esempio alla Statistica, sono di grande importanza le medie aritmetiche di insiemi di numeri. È noto, per esempio, che, se si fanno n osservazioni (misure) di una certa quantità ottenendo i risultati $x_1, x_2, \ldots, x_n$, si usa assumere, come valore della quantità osservata, la media aritmetica $\overline{x} = (x_1 + x_2 + \ldots + x_n)/n$. Il Calcolo delle Probabilità fornisce buone giustificazioni per questo modo di procedere, facendo appello ad alcuni "teoremi limite" che trattano appunto del limite di successioni di medie aritmetiche.

In questo paragrafo ci limiteremo a dare qualche definizione e mostrare qualche risultato relativi alle medie aritmetiche.

Sia $\{a_n\}$ una successione a valori reali e sia $\{\alpha_n\}$ la successione delle medie aritmetiche:

$$\alpha_n = \frac{a_0 + a_1 + \ldots + a_n}{n+1} .$$

Se $\{a_n\}$ è convergente, cioè

$$\lim_{n\to+\infty} a_n = l \in \mathbb{R}$$

abbiamo visto (o, meglio, abbiamo proposto, nell'Esercizio 4.3.13 di dimostrare) che anche la successione $\{\alpha_n\}$ è convergente allo stesso limite:

$$\lim_{n\to+\infty} \frac{a_0 + a_1 + \ldots + a_n}{n+1} = l .$$

Si prova inoltre che, se $\{a_n\}$ diverge, anche $\{\alpha_n\}$ diverge.

Ma ci sono successioni indeterminate che, passando alle medie, vengono trasformate in successioni convergenti. Per esempio, sia $\{a_n\}$ la successione

$$+1, -1, +1, -1, +1, -1, \ldots$$

Allora la corrispondente $\{\alpha_n\}$ è

$$1, 0, \frac{1}{3}, 0, \frac{1}{5}, 0, \frac{1}{7}, \ldots$$

la quale è convergente a zero.

Perciò il passaggio dalla successione $\{a_n\}$ alla successione $\{\alpha_n\}$ consente di estendere l'operazione di limite; parleremo di *limite alla Cesáro* per indicare il limite della successione delle medie.

Passiamo al caso delle serie. Data la successione $\{a_n\}$, consideriamo la successione delle somme parziali $\{s_n\}$; se $\{s_n\}$ converge (o diverge), la serie $\sum a_n$ è detta convergente (divergente); se $\{s_n\}$ è indeterminata, possiamo passare alla successione delle medie σ_n:

$$\sigma_n = \frac{s_0 + s_1 + \ldots + s_n}{n+1}$$

e vedere se questa converge; se ciò accade diremo che la serie $\sum a_n$ è *sommabile secondo Cesáro* (o, nel senso delle medie aritmetiche).

Per esempio, la serie

$$\sum_{n=0}^{\infty}(-1)^n = 1 - 1 + 1 - 1 + \ldots$$

è indeterminata. La successione delle sue somme parziali, $\{s_n\}$, è

$$1, 0, 1, 0, 1, 0, 1, \ldots$$

Ma la succesione $\{\sigma_n\}$ è

$$1, \frac{1}{2}, \frac{2}{3}, \frac{1}{2}, \frac{3}{5}, \frac{1}{2}, \frac{4}{7}, \ldots$$

la quale converge a $\frac{1}{2}$. Diremo perciò che la serie $\sum_{n=0}^{\infty}(-1)^n$ è sommabile (alla Cesáro) con somma $\frac{1}{2}$.

Abbiamo così ampliato la classe delle serie convergenti. Naturalmente ciò ha un prezzo: per le serie convergenti alla Cesáro non vale più nemmeno la proprietà associativa.

Dunque, tenendo conto dei risultati visti nei paragrafi 2.4 e 2.5, possiamo dire che le serie assolutamente convergenti sono quelle che, in un certo senso, possono essere trattate come somme finite: per esse infatti valgono le usuali proprietà associativa e commutativa, e su di esse si possono eseguire le operazioni aritmetiche con le solite regole. Alcune di queste proprietà vengono meno passando alle serie convergenti (in senso ordinario) e altre ancora si perdono passando alle serie sommabili alla Cesáro.

Vogliamo, in conclusione, ricordare (senza dimostrazione) il seguente risultato (dovuto appunto a Cesáro) che estende il teorema di Mertens sul prodotto di due serie.

Proposizione 2.19 - *Siano $\sum a_n$ e $\sum b_n$ due serie convergenti con somma A e B rispettivamente. Allora la serie $\sum c_n$, prodotto alla Cauchy delle due serie date, è sommabile alla Cesáro con somma AB.*

Esercizi

1. Discutere la convergenza delle seguenti serie

$$\sum_{n=1}^{\infty} \frac{1}{n+\sqrt{n}} \,,\quad \sum_{n=1}^{\infty} n^{-\log n} \,,\quad \sum_{n=2}^{\infty} \frac{1}{\log(n!)}$$

$$\sum_{n=2}^{\infty} \frac{1}{\sqrt{n^3-n}} \,,\quad \sum_{n=1}^{\infty} \frac{n+\log n}{(n+\cos n)^3} \,,\quad \sum_{n=2}^{\infty} \left(\frac{\log\log n}{\log n}\right)^n .$$

2. Discutere, al variare dei parametri reali $x, a, b, \ldots$ la convergenza, e la convergenza assoluta, delle seguenti serie

$$\sum_{n=1}^{\infty} \frac{1}{1+x^n} \,,\quad \sum_{n=k}^{\infty} \frac{n^a}{\binom{n}{k}} \quad,\quad \sum_{n=1}^{\infty} [\sin(a+\frac{b}{n})]^n \qquad 0 \le a < \frac{\pi}{2} .$$

3. Mostrare, usando il criterio di condensazione, che la serie

$$\sum_{n=2}^{\infty} \frac{1}{n^\alpha (\log n)^\beta} \quad (\alpha, \beta \in \mathbb{R})$$

è convergente per $\alpha > 1, \beta \in \mathbb{R}$ e per $\alpha = 1, \beta > 1$; divergente negli altri casi.

4. Trovare per quali valori di x reale risultano convergenti (convergenti assolutamente) le seguenti serie:

$$a)\ \sum_{n=0}^{\infty} \frac{x^n}{n!} \quad b)\ \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n}}{(2n)!} \quad c)\ \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!} .$$

Mostrare, come si è fatto nell'Esempio 2.10 in un caso particolare, che la somma delle serie $a), b), c)$ è: $e^x, \cos x, \sin x,\ \forall\ x \in \mathbb{R}$ rispettivamente.

5. Usando una calcolatrice tascabile, calcolare, con un errore inferiore a $1/1000$, il valore di

$$\frac{1}{\sqrt{e}} \,,\ \sin\frac{\pi}{5} \,,\ \cos 1$$

(usare l'esercizio 4 e il teorema 2.12).

6. Trovare per quali valori di x reale risultano convergenti (assolutamente convergenti) le serie:

$$a)\ \sum_{n=1}^{\infty} (-1)^{n+1} \frac{x^n}{n} \quad b)\ \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{2n+1} .$$

Esaminare in particolare i valori $x = 1$ e $x = -1$.

7. Mostrare, utilizzando gli sviluppi di MacLaurin delle funzioni $\log(1+x)$ e $\operatorname{arctg} x$, con il resto nella forma di Lagrange, che, per $-1 < x < 1$, la somma delle serie $a), b)$ dell'esercizio 6, è $\log(1+x)$ e $\operatorname{arctg} x$ rispettivamente.

Dalle formule trovate nell'esercizio 1.14 si ricava

$$\frac{2\cdot 2\cdot 4\cdot 4\cdot 6\cdot 6\cdot \ldots \cdot 2p\cdot 2p}{3\cdot 3\cdot 5\cdot 5\cdot \ldots \cdot (2p-1)(2p-1)(2p+1)} < \frac{\pi}{2}$$

$$\frac{\pi}{2} < \frac{2\cdot 2\cdot 4\cdot 4\cdot 6\cdot 6\cdot \ldots \cdot (2p-2)(2p-2)2p}{3\cdot 3\cdot 5\cdot 5\cdot \ldots \cdot (2p-1)(2p-1)} .$$

Posto $W_p = \left(\frac{2\cdot 4\cdot 6\cdot \ldots \cdot 2p}{3\cdot 5\cdot \ldots \cdot (2p-1)}\right)^2 \frac{1}{2p+1}$

si ha $W_p < \frac{\pi}{2} < W_p \frac{2p+1}{2p}$

Si ha perciò: $\lim_{p\to+\infty} \frac{\pi}{2}\frac{1}{W_p} = 1$, cioè

$$\frac{\pi}{2} = \lim_{p\to+\infty} W_p .$$

17. *Riordinamenti.* Consideriamo la serie (vedi esercizio 7)

$$1-\frac{1}{2}+\frac{1}{3}-\frac{1}{4}+\frac{1}{5}-\frac{1}{6}+\frac{1}{7}-\frac{1}{8}+\frac{1}{9}-\frac{1}{10}+\ldots = \log 2 .$$

Moltiplichiamo i termini per $\frac{1}{2}$

$$\frac{1}{2} - \frac{1}{4} + \frac{1}{6} - \frac{1}{8} + \frac{1}{10} - \ldots = \frac{1}{2}\log 2 .$$

Sommando le due serie otteniamo:

$$1+\frac{1}{3}-\frac{1}{2}+\frac{1}{5}+\frac{1}{7}-\frac{1}{4}+\frac{1}{9}+\frac{1}{11}-\frac{1}{6}+\ldots = \frac{3}{2}\log 2 .$$

Si noterà che l'ultima serie è un riordinamento della prima, ma la sua somma è diversa. Spiegarne la ragione.

3. ESTENSIONI DELL'INTEGRALE DI RIEMANN

3.1 Integrali impropri

Cercheremo ora di estendere l'operazione di integrazione a funzioni non limitate e ad intervalli non limitati. Procediamo per passi successivi.

Cominciamo col considerare funzioni non necessariamente limitate su un intervallo limitato $[a,b]$.

Sia $f:(a,b] \to \mathbb{R}$ tale che, $\forall\, \delta > 0$, $f \in \mathcal{R}(a+\delta, b)$, cioè esiste l'integrale

$$I(\delta) = \int_{a+\delta}^{b} f(x)dx .$$

Definizione 3.1 - *Se esiste finito il limite*

$$\lim_{\delta \to 0_+} I(\delta) \tag{3.1}$$

la funzione f si dirà integrabile in senso improprio nell'intervallo $[a,b]$; il limite (3.1) *si chiamerà integrale (generalizzato o improprio) di f in $[a,b]$ e si denoterà ancora con il simbolo*

$$\int_a^b f(x)dx \ .$$

Questa definizione è una estensione di quella data in 1.1 perché, se $f \in \mathcal{R}(a,b)$, il limite (3.1) esiste (per la continuità della funzione integrale) ed è uguale all'integrale di f in $[a,b]$.

***Esempio* 3.1** - Sia $f(x) = x^{-\alpha}$ $(\alpha > 0)$ e $[a,b]$ sia l'intervallo $[0,1]$. La funzione integranda è definita e continua su $(0,1]$, ma non limitata; essa è però limitata in $[\delta,1]$ $\forall\ \delta > 0$ e risulta

$$I(\delta) = \int_\delta^1 x^{-\alpha}dx = \begin{cases} \dfrac{1}{1-\alpha}(1-\delta^{1-\alpha}) & \text{se } \alpha \neq 1 \\ -\log\delta & \text{se } \alpha = 1 \end{cases}$$

Perciò si ha:

$$\lim_{\delta \to 0_+} I(\delta) = \begin{cases} \dfrac{1}{1-\alpha} & \text{se } \alpha < 1 \\ +\infty & \text{se } \alpha \geq 1 \end{cases}$$

La funzione $x^{-\alpha}$ è dunque integrabile in senso improprio sull'intevallo $[0,1]$ se $\alpha < 1$ e si ha:

$$\int_0^1 x^{-\alpha}dx = \frac{1}{1-\alpha}$$

(il risultato è ovviamente vero anche se $\alpha \leq 0$); l'integrale non esiste (o, come si dice anche, non converge) se $\alpha \geq 1$.

La definizione 3.1 si adatta subito al caso in cui $f : [a,b) \to \mathbb{R}$ non sia limitata nell'intorno dell'estremo destro b, ma risulti integrabile in $[a, b-\delta]$ $\forall\delta > 0$; porremo infatti

$$\int_a^b f(x)dx := \lim_{\delta \to 0_+} \int_a^{b-\delta} f(x)dx$$

se tale limite esiste finito.

Se la funzione f non è limitata nell'intorno di un punto c interno ad $[a, b]$ diremo che è integrabile in senso improprio se lo è in $[a, c]$ e in $[c, b]$ e porremo

$$\int_a^b f(x)dx := \int_a^c f(x)dx + \int_c^b f(x)dx \tag{3.2}$$

Se poi f non è limitata nell'intorno dei punti $x_1, x_2, \ldots, x_n \in [a, b]$, con $a < x_1 < x_2 < \ldots < x_n < b$ divideremo l'intervallo $[a, b]$ in tanti sottointervalli in modo che in ognuno di questi la funzione f sia illimitata al più in uno degli estremi; f si dirà integrabile in senso improprio su $[a, b]$ se lo è in ciascuno di questi intervallini (si riconosce subito che l'integrabilità, come il valore dell'integrale, non dipendono dalla suddivisione scelta).

Esempio **3.2** - Sia dato l'integrale

$$\int_{-1}^{+1} \frac{dx}{\sqrt{1-x^2}} \, .$$

La funzione integranda non è limitata in entrambi gli estremi dell'intervallo $[-1, +1]$. Possiamo studiare la convergenza dell'integrale considerando separatamente i due integrali $\int_{-1}^{0}$ e $\int_{0}^{1}$. Possiamo anche considerare direttamente l'integrale

$$I(\delta_1, \delta_2) = \int_{\delta_1}^{\delta_2} \frac{dx}{\sqrt{1-x^2}} = [\arcsin x]_{\delta_1}^{\delta_2} = \arcsin \delta_2 - \arcsin \delta_1$$

e quindi prendere il limite

$$\lim_{\substack{\delta_1 \to -1_+ \\ \delta_2 \to 1_-}} I(\delta_1, \delta_2) = \frac{\pi}{2} - \left(-\frac{\pi}{2}\right) = \pi \, .$$

Si osservi che il limite va preso facendo tendere δ_1 a -1_+ e δ_2 a 1_- in maniera indipendente uno dall'altro.

Consideriamo ora intervalli non limitati.

Sia $f : [a, +\infty) \to \mathbb{R}$ integrabile su ogni intervallo $[a, \omega]$ con $\omega > a$. Poniamo

$$J(\omega) = \int_a^\omega f(x)dx \, .$$

Definizione 3.2 - *Se esiste finito il limite*

$$\lim_{\omega \to +\infty} J(\omega) \tag{3.3}$$

la funzione f si dirà integrabile in senso improprio sull'intervallo $[a,+\infty)$; il limite (3.3) *si dirà integrale generalizzato o improprio di f in $[a,+\infty)$ e si indicherà col simbolo*

$$\int_a^{+\infty} f(x)dx \, .$$

Analoga è la definizione dell'integrale improprio

$$\int_{-\infty}^{b} f(x)dx$$

di una funzione $f : (-\infty, b] \to \mathbb{R}$, integrabile su ogni intervallo $[-\omega, b]$ con $-\omega \leq b$.

Per l'integrale da $-\infty$ a $+\infty$ di una funzione $f : \mathbb{R} \to \mathbb{R}$, integrabile su ogni intervallo limitato, si pone

$$\int_{-\infty}^{+\infty} f(x)dx := \int_{-\infty}^{c} f(x)dx + \int_{c}^{+\infty} f(x)dx \tag{3.4}$$

se c è un qualsiasi numero reale e gli integrali impropri a secondo membro della (3.4) sono convergenti. Ancora si noterà che la definizione (3.4) non dipende dalla scelta di c.

Esempi

3.3. Sia $f(x) = x^{-\alpha}$. Calcoliamo, per $\omega \geq 1$

$$J(\omega) = \int_1^{\omega} x^{-\alpha} dx = \begin{cases} \dfrac{1}{1-\alpha}(\omega^{1-\alpha} - 1) & \text{se } \alpha \neq 1 \\ \log \omega & \text{se } \alpha = 1 \end{cases}$$

Si ha allora

$$\lim_{\omega \to +\infty} J(\omega) = \begin{cases} \dfrac{1}{\alpha - 1} & \text{se } \alpha > 1 \\ +\infty & \text{se } \alpha \leq 1 \end{cases}$$

La funzione $x^{-\alpha}$ è dunque integrabile in senso improprio sulla semiretta $[1, +\infty)$ se $\alpha > 1$ e si ha

$$\int_1^{+\infty} x^{-\alpha} dx = \frac{1}{\alpha - 1} \, .$$

L'integrale non esiste (non converge) se $\alpha \leq 1$.

3.4. $\displaystyle\int_0^{\omega} e^{-x} dx = [-e^{-x}]_0^{\omega} = 1 - e^{-\omega} \to 1$ per $\omega \to +\infty$.

Perciò abbiamo $\int_0^{+\infty} e^{-x}dx = 1$.

3.5. Per calcolare $\int_{-\infty}^{+\infty} \frac{dx}{1+x^2}$ possiamo calcolare direttamente

$$\lim_{\substack{\omega_1 \to -\infty \\ \omega_2 \to +\infty}} \int_{\omega_1}^{\omega_2} \frac{dx}{1+x^2} = \lim_{\substack{\omega_1 \to -\infty \\ \omega_2 \to +\infty}} [\operatorname{arctg} x]_{\omega_1}^{\omega_2} =$$

$$= \lim_{\omega_2 \to +\infty} \operatorname{arctg} \omega_2 - \lim_{\omega_1 \to -\infty} \operatorname{arctg} \omega_1 = \frac{\pi}{2} - \left(-\frac{\pi}{2}\right) = \pi \, .$$

Molte delle proprietà classiche degli integrali restano valide per gli integrali generalizzati. In particolare valgono le proprietà di omogeneità, additività (sia rispetto alla funzione integranda che rispetto all'intervallo di integrazione); vale il teorema della media (esercizio 2), se l'intervallo di integrazione è limitato e l'integranda è continua nei punti interni di tale intervallo; anche il teorema fondamentale del calcolo si estende agli integrali impropri (vedi esercizio 3).

Una proprietà fondamentale che invece viene a cadere è l'integrabilità di $|f|$ se f è integrabile; ma su questo punto torneremo più avanti.

3.2 Criteri di convergenza

Sarà utile naturalmente disporre di criteri che garantiscano la convergenza degli integrali impropri.

■ **Teorema 3.1** - (Criterio del confronto).

1. *Siano f, g due funzioni definite in $[a, +\infty)$ e integrabili in ogni intervallo limitato $[a, \omega]$ con $\omega \geq a$; esista $x_0 \geq a$ tale che*

$$0 \leq f(x) \leq g(x) \qquad \textit{per} \quad x \geq x_0 \, .$$

Allora, se g è integrabile in $[a, +\infty)$, anche f lo è.

2. *Siano f, g due funzioni definite in $(a, b]$ e integrabili in ogni intervallo $[a+\delta, b]$ con $\delta > 0$; esista $x_0 \in (a, b]$ tale che:*

$$0 \leq f(x) \leq g(x) \qquad \textit{per} \quad a < x \leq x_0 \, .$$

Allora, se g è integrabile in $[a, b]$, anche f lo è.

Dimostrazione di 1. – La funzione integrale $J(\omega) = \int_a^\omega f(x)dx$ è crescente (almeno per $\omega \geq x_0$); perciò esiste $\lim_{\omega \to +\infty} J(\omega)$; bisogna mostrare che è finito.

D'altra parte g è positiva (per $x \geq x_0$) e integrabile su $[a, +\infty)$; abbiamo:

$$\begin{aligned} J(\omega) &= \int_a^{x_0} f(x)dx + \int_{x_0}^{\omega} f(x)dx \leq \\ &\leq \int_a^{x_0} f(x)dx + \int_{x_0}^{\omega} g(x)dx \leq \\ &\leq \int_a^{x_0} f(x)dx + \int_{x_0}^{+\infty} g(x)dx < +\infty \,. \end{aligned}$$

Dunque $J(\omega)$ è limitata, perciò ammette limite finito. Analoga è la dimostrzione di 2. □

Per mezzo del criterio del confronto, e utilizzando il risultato degli esempi 1. e 2. si dimostrano facilmente le seguenti affermazioni.

Sia $f : [a, +\infty) \to \mathbb{R}$ una funzione non negativa e integrabile su ogni intervallo limitato $[a, \omega]$ con $\omega \geq a$.

1. *Condizione sufficiente perché esista $\int_a^{+\infty} f(x)dx$ è che $f(x)$ sia infinitesima, per $x \to +\infty$, di ordine α rispetto a $\frac{1}{x}$, con $\alpha > 1$.*

2. *Se $f(x) \to L \neq 0$ per $x \to +\infty$, oppure è infinitesima di ordine ≤ 1 rispetto all'infinitesimo campione $\frac{1}{x}$, l'integrale $\int_a^{+\infty} f(x)dx$ non converge.*

Sia $f : (a, b] \to \mathbb{R}$ una funzione non negativa e integrabile su ogni intervallo $[a + \delta, b]$ con $\delta > 0$.

3. *Condizione sufficiente perché esista $\int_a^b f(x)dx$ è che $f(x)$ risulti infinita per $x \to a_+$ di ordine al più α, con $\alpha < 1$, rispetto all'infinito campione $\frac{1}{x - a}$.*

4. *Se $f(x)$ è infinita per $x \to a_+$ di ordine ≥ 1, l'integrale $\int_a^b f(x)dx$ non converge.*

I criteri sopra enunciati si estendono in modo ovvio al caso in cui f sia definita sull'intervallo $(-\infty, b]$ oppure $[a, b)$.

Esempi

3.6. Gli integrali (con integrande positive)

$$\int_0^{\infty} x^{\alpha} e^{-x} dx \ \ (\alpha \geq 0) \ , \ \int_1^{\infty} \frac{1}{\sqrt{x}} \log\left(1 + \frac{1}{x}\right) \ , \ \int_1^{\infty} \left(\sin \frac{1}{x}\right)^2 dx$$

sono convergenti. Infatti, preso come infinitesimo campione $\frac{1}{x}$ per $x \to +\infty$ abbiamo:

$x^\alpha e^{-x}$, $\forall\alpha \geq 0$, è infinitesima di ordine superiore ad ogni numero fissato.

$\frac{1}{\sqrt{x}}\log\left(1+\frac{1}{x}\right) \sim \left(\frac{1}{x^{3/2}}\right)$ è infinitesima di ordine 3/2

$\left(\sin\frac{1}{x}\right)^2 \sim \left(\frac{1}{x^2}\right)$ è infinitesima di ordine 2.

3.7. L'integrale $\int_e^{+\infty}\frac{dx}{x\log x}$ non converge. Infatti

$$\int_e^\omega \frac{dx}{x\log x} = [\log\ \log x]_e^\omega = \log\ \log\omega \to +\infty \text{ per } \omega \to +\infty .$$

Si osservi tuttavia che la funzione integranda (positiva) è infinitesima, per $x \to +\infty$, di ordine superiore a $\frac{1}{x}$; ma non esiste alcun α, $\alpha > 1$, per cui si possa affermare che l'integranda è infinitesima di ordine α.

3.8. Gli integrali (con integrande positive)

$$\int_0^{\pi/2}\frac{dx}{\sqrt{\operatorname{tg} x}} \quad , \quad \int_0^1 \frac{dx}{e^{\sqrt{x}}-1}$$

sono convergenti. Infatti, per $x \to 0$, entrambe le integrande sono infiniti di ordine $\frac{1}{2}$ rispetto a $\frac{1}{x}$. (Per $x \to \pi/2$ la prima integranda tende a zero).

Invece gli integrali

$$\int_0^1 \exp\left(-\frac{1}{x-1}\right)dx \quad \int_0^1 \frac{x-1}{\log(1-x)}dx$$

non esistono. Infatti $\exp\left(-\frac{1}{x-1}\right) \to +\infty$ per $x \to 1_-$ più rapidamente di qualunque potenza $\frac{1}{(x-1)^\alpha}$; $\frac{x-1}{\log(1-x)}$ è infinito dello stesso ordine di $\frac{1}{x}$ per $x \to 0_+$ (mentre tende a zero per x tendente a 1_-).

Se la funzione $f(x)$ non è sempre dello stesso segno, possiamo ancora applicare il teorema 3.1 alla funzione $|f(x)|$ per decidere circa l'integrabilità di $f(x)$. Infatti vale il seguente

▪ **Teorema 3.2** - *Sia* $f : [a,+\infty) \to \mathbb{R}$, *integrabile in* $[a,\omega]$, $\forall\omega \in \mathbb{R}$, *e* $|f(x)|$ *sia integrabile su* $[a,+\infty)$; *allora anche* f *lo è e risulta*

$$\left|\int_a^{+\infty} f(x)dx\right| \leq \int_a^{+\infty} |f(x)|dx .$$

Analogamente, se $f : [a, b] \to \mathbb{R}$ *è integrabile in* $[a+\delta, b]$, $\forall \delta > 0$, *e tale che* $|f(x)|$ *è integrabile su* $[a, b]$, *allora anche* f *lo è e risulta*

$$\left|\int_a^b f(x)dx\right| \leq \int_a^b |f(x)|dx \ .$$

Dimostrazione - Poiché f è integrabile in $[a, \omega]$ $\forall \omega \geq a$, tali risultano anche $|f|$, f_+, f_- (proprietà 3 del teorema 1.8). Poiché abbiamo

$$0 \leq f_+(x) \leq |f(x)| \ , \ 0 \leq f_-(x) \leq |f(x)|$$

ne consegue, se $|f|$ è integrabile su $[a, +\infty)$, per il teorema del confronto, che anche f_+ e f_- e perciò $f = f_+ - f_-$ sono integrabili su $[a, +\infty)$. Risulta anche

$$\begin{aligned}\left|\int_a^{+\infty} f(x)dx\right| &= \left|\int_a^{+\infty} f_+(x)dx - \int_a^{+\infty} f_-(x)dx\right| \leq \\ &\leq \int_a^{+\infty} f_+(x)dx + \int_a^{+\infty} f_-(x)dx = \\ &= \int_a^{+\infty} |f(x)|dx\end{aligned}$$

Analoga è la dimostrazione nel caso dell'intervallo $[a, b]$. □

È importante notare che, al contrario di quanto accade per l'integrale di Riemann, l'integrabilità di f non implica quella di $|f|$. Come le serie, gli integrali impropri possono essere convergenti, ma non assolutamente convergenti, cioè può esistere $\int f$ ma non $\int |f|$.

***Esempio* 3.9** - L'integrale $\displaystyle\int_0^{+\infty} \frac{\sin x}{x} dx$ è convergente, ma non assolutamente convergente.

Basterà evidentemente considerare l'integrale da $\pi/2$ a $+\infty$. Con una integrazione per parti abbiamo:

$$\begin{aligned}\int_{\pi/2}^{\omega} \frac{\sin x}{x} dx &= \left[-\frac{\cos x}{x}\right]_{\pi/2}^{\omega} - \int_{\pi/2}^{\omega} \frac{\cos x}{x^2} dx = \\ &= -\frac{\cos \omega}{\omega} - \int_{\pi/2}^{\omega} \frac{\cos x}{x^2} dx \ .\end{aligned}$$

L'ultimo integrale esiste, essendo $\dfrac{|\cos x|}{x^2} \leq \dfrac{1}{x^2}$. Prendendo il limite per $\omega \to +\infty$ abbiamo allora

$$\int_{\pi/2}^{+\infty} \frac{\sin x}{x} dx = -\int_{\pi/2}^{+\infty} \frac{\cos x}{x^2} dx \ .$$

Per mostrare che l'integrale proposto non è assolutamente convergente osserviamo che, per n intero positivo, abbiamo:

$$\int_0^{n\pi} \frac{|\sin x|}{x} dx = \sum_{k=1}^{n} \int_{(k-1)\pi}^{k\pi} \frac{|\sin x|}{x} dx > \sum_{k=1}^{n} \frac{1}{k\pi} \int_{(k-1)\pi}^{k\pi} |\sin x| dx =$$

$$= \frac{2}{\pi} \sum_{k=1}^{n} \frac{1}{k} .$$

Poiché la serie armonica diverge, si ottiene

$$\lim_{n\to+\infty} \int_0^{n\pi} \frac{|\sin x|}{x} dx = +\infty .$$

3.3 Serie e integrali

Le serie possono essere considerate come particolari integrali impropri su una semiretta. Infatti, sia data la serie $\sum_{n=0}^{\infty} a_n$ e sia $\{A_n\}$ la successione delle ridotte. Definiamo una funzione a scala

$$f(x) = a_k \quad \text{per } k \leq x < k+1 \ , \ k = 0,1,2,\dots$$

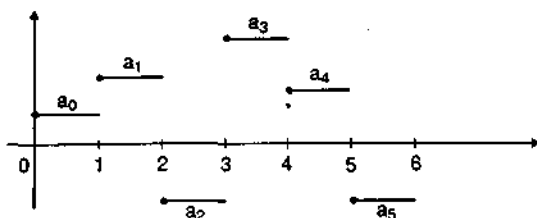



Fig. 8.8

Risulta allora

$$A_n = \sum_{k=0}^{n} a_k = \int_0^{n+1} f(x)dx .$$

Perciò se l'integrale improprio $\int_0^{\infty} f(x)dx$ esiste, allora la serie converge e risulta

$$\sum_{n=0}^{\infty} a_n = \int_0^{\infty} f(x)dx . \tag{3.5}$$

È vero anche il viceversa, poiché, $\forall \omega > 0$, si ha:

$$\int_0^\omega f(x)dx = \int_0^{[\omega]} f(x)dx + (\omega)a_{[\omega]} = \sum_{k=0}^{[\omega]-1} a_k + (\omega)a_{[\omega]} \,.$$

Perciò, se la serie converge, esiste finito $\lim\limits_{\omega \to +\infty} \int_0^\omega f(x)dx$ e vale la (3.5).

La seguente proposizione ci fornisce un utile criterio di convergenza per le serie mediante confronto con un integrale.

Proposizione 3.3 - *Sia data la serie* $\sum\limits_{n=0}^{\infty} a_n$. *Se esiste una funzione non negativa* $f : [0, +\infty) \to \mathbb{R}$ *integrabile su* $[0, +\infty)$ *e tale che*

$$|a_n| \le f(x) \;\; \text{per } n \le x \le n+1 \;,\; n \in \mathbb{N} \,.$$

allora la serie è assolutamente convergente.

Dimostrazione - Posto $c_n = \int_n^{n+1} f(x)dx$, $n \in \mathbb{N}$, la serie $\sum\limits_{n=0}^{\infty} c_n$, per quanto si è detto nell'introduzione al paragrafo, converge e la sua somma è uguale a $\int_0^\infty f(x)dx$. Ma

$$|a_n| = \int_n^{n+1} |a_n| dx \le \int_n^{n+1} f(x)dx = c_n \,.$$

Pertanto la serie $\sum |a_n|$ converge per il criterio del confronto. □

***Esempio* 3.10** - La convergenza della serie $\sum\limits_{n=0}^{\infty} \frac{1}{(n+1)^\alpha}$ per $\alpha > 1$ può essere stabilita per confronto con l'integrale $\int_1^\infty \frac{dx}{x^\alpha}$; infatti l'integrale converge per $\alpha > 1$ e inoltre si ha

$$\frac{1}{x^\alpha} \ge \frac{1}{(n+1)^\alpha} \quad \text{se} \quad n \le x \le n+1 \;,\; n \in \mathbb{N} \,.$$

*3.4 Integrale di Stieltjes

Il problema geometrico del calcolo dell'area di una figura piana (per esempio, un trapezioide) porta a considerare somme del tipo $\sum\limits_{i=1}^{n} f(x_i)\Delta x_i$. In svariate questioni

della matematica (per esempio, teoria degli operatori) e delle scienze applicate (per esempio, Statistica e Calcolo delle probabilità) si incontrano problemi che portano alla considerazione di somme del tipo

$$\sum_{i=1}^{n} f(x_i)[g(x_{i+1}) - g(x_i)]$$

che possono essere considerate una generalizzazione delle somme integrali alla Riemann, ove l'incremento $\Delta x_i = x_{i+1} - x_i$ della variabile indipendente è stato sostituito dall'incremento $g(x_{i+1}) - g(x_i)$ di una funzione g. Sulla funzione g faremo l'ipotesi che sia monotona (per esempio, crescente).

Siano dunque f e g due funzioni: $[a,b] \to \mathbb{R}$, f limitata e g monotona. Fatta una suddivisione $\mathcal{D}(x_0, x_1, \ldots, x_n)$ dell'intervallo $[a,b]$, consideriamo le somme

$$s = s(\mathcal{D}, f, g) = \sum_{i=1}^{n} m_i[g(x_i) - g(x_{i-1})]$$

$$S = S(\mathcal{D}, f, g) = \sum_{i=1}^{n} M_i[g(x_i) - g(x_{i-1})]$$

avendo posto, come al solito

$$m_i = \inf_{(x_{i-1}, x_i)} f \qquad M_i = \sup_{(x_{i-1}, x_i)} f \ .$$

Esse si diranno rispettivamente *somma inferiore e superiore* di Stieltjes. Come si è fatto per le analoghe somme di Riemann si mostra che, per una data coppia di funzioni f, g, ogni somma inferiore è $\leq$ ogni somma superiore.

Definizione 3.3 - *Diciamo che la funzione (limitata) $f : [a,b] \to \mathbb{R}$ è integrabile rispetto alla funzione (monotona) $g : [a,b] \to \mathbb{R}$ quando l'estremo superiore delle precedenti somme inferiori coincide con l'estremo inferiore delle somme superiori. Il valore comune dei due estremi è detto* **integrale di Stieltjes** *di f (funzione integranda) rispetto a g (funzione integratrice) e indicato con il simbolo*

$$\int_a^b f\,dg \quad \textit{o anche} \quad \int_a^b f(x)dg(x) \ .$$

Osservazione 3.1 - Il simbolo dg nella precedente scrittura non va inteso come differenziale della funzione g; questa infatti, essendo monotona, può benissimo non essere differenziabile. Si osserverà anche che, se $g(x) = x$, l'integrale sopra definito coincide con l'integrale classico di Riemann; abbiamo perciò ottenuto una estensione di quest'ultimo.

Sulla traccia delle considerazioni svolte nella sezione 1, lo studente può senza difficoltà dimostrare la seguente

Proposizione 3.4 - *Condizione necessaria e sufficiente perché una funzione limitata f sia integrabile su $[a,b]$ rispetto alla funzione monotona g è che, per ogni $\varepsilon > 0$, sia possibile trovare una suddivisione $\mathcal{D}$ di $[a,b]$ per la quale risulti*

$$\mathcal{S} - s < \varepsilon .$$

Si pone subito la questione di individuare classi di funzioni integrabili; i seguenti due teoremi forniscono una prima risposta.

■ **Teorema 3.5** - *Se f è continua su $[a,b]$ e g è monotona, allora l'integrale $\int_a^b f\,dg$ esiste.*

Dimostrazione - Fissato $\varepsilon > 0$, per l'uniforme continuità di f in $[a,b]$, si può trovare δ tale che

$$|f(x') - f(x'')| < \varepsilon \quad \text{se} \quad |x' - x''| < \delta .$$

Allora consideriamo una suddivisione $\mathcal{D}$ di $[a,b]$ di ampiezza inferiore a δ; abbiamo (supponendo g crescente):

$$\mathcal{S} - s = \sum_{i=1}^{n}(M_i - m_i)[g(x_i) - g(x_{i-1})] <$$
$$< \varepsilon \sum_{i=1}^{n}[g(x_i) - g(x_{i-1})] = \varepsilon[g(b) - g(a)] .$$

Per l'arbitrarietà di ε, segue la tesi. □

Un caso particolare di questo teorema, molto importante per le applicazioni, si ha quando g è una funzione a scala. Siano dunque $\xi_1, \xi_2, \ldots, \xi_m$ i punti di discontinuità della funzione g e $p_1, p_2, \ldots, p_m$ i rispettivi salti. Presa una qualsiasi suddivisione di $[a,b]$ i termini $g(x_i) - g(x_{i-1})$ sono tutti nulli tranne quelli per cui nell'intervallino $[x_{i-1}, x_i]$ viene a cadere uno ξ_k (se la suddivisione è abbastanza fine, nessun intervallino può contenere più di uno ξ_k); perciò le somme superiori $\mathcal{S}$ e le somme inferiori s conterranno soltanto precisamente m addendi; si ottiene, per la continuità di f,

$$\int_a^b f\,dg = \sum_{k=1}^{m} f(\xi_k)p_k$$

cioè l'integrale è, in questo caso, una somma finita!

Se, nelle stesse ipotesi sopra enunciate, g ha una infinità numerabile di salti, si ottiene

$$\int_a^b f\,dg = \sum_{k=1}^{\infty} f(\xi_k)p_k .$$

■ **Teorema 3.6** - *Se f è monotona e g è (monotona e) continua, l'integrale $\int_a^b f dg$ esiste.*

Dimostrazione - Per la continuità di g possiamo considerare una suddivisione $\mathcal{D}(x_0, \ldots, x_n)$ di $[a, b]$ per cui risulti

$$g(x_i) - g(x_{i-1}) = \frac{g(b) - g(a)}{n} .$$

Per una tale suddivisione avremo (per fissare le idee, sia f crescente):

$$\begin{aligned} S - s &= \sum_{i=1}^{n}(M_i - m_i)[g(x_i) - g(x_{i-1})] = \\ &= \frac{g(b) - g(a)}{n} \sum_{i=1}^{n}[f(x_i) - f(x_{i-1})] = \\ &= \frac{1}{n}[g(b) - g(a)]\,[f(b) - f(a)] . \end{aligned}$$

Fissato $\varepsilon > 0$ potremo prendere n abbastanza grande perché risulti $S - s < \varepsilon$. □

Se nel teorema 3.5 f non è continua, oppure, nel teorema 3.6, g non è continua, l'integrale può non esistere.

***Esempio* 3.11** - Sull'intervallo $[1, 4]$ consideriamo le funzioni

$$f(x) = g(x) = \begin{cases} -1 & \text{se } x \in [1, 2) \\ +1 & \text{se } x \in [2, 4] . \end{cases}$$

Per una qualsiasi suddivisione di $[1, 4]$, il termine $g(x_i) - g(x_{i-1})$ è sempre nullo tranne che quando l'intervallino $[x_{i-1}, x_i]$ contenga il punto $x = 2$. In questo caso abbiamo:

$$s = -2 \qquad S = +2 .$$

Pertanto f non è integrabile rispetto a sé stessa.

Per l'integrale di Stiltjes valgono proprietà analoghe a quelle già illustrate per l'integrale di Riemann, che qui sotto elenchiamo senza dimostrazione. Supporremo le funzioni f, f_1, f_2 integrabili rispetto alle funzioni integratrici g, g_1, g_2 monotone crescenti:

a) $\int_a^b (c_1 f_1 + c_2 f_2) dg = c_1 \int_a^b f_1 dg + c_2 \int_a^b f_2 dg$
(linearità rispetto alla funzione integranda)

b) $\int_a^b f d(c_1 g_1 + c_2 g_2) = c_1 \int_a^b f dg_1 + c_2 \int_a^b f dg_2$

(linearità rispetto alla funzione integratrice)

c) $\int_a^b f dg = \int_a^c f dg + \int_c^b f dg \qquad (a < c < b)$

(additività rispetto all'intervallo di integrazione)

d) $f_1 \leq f_2 \Rightarrow \int_a^b f_1 dg \leq \int_a^b f_2 dg$

e) $\left| \int_a^b f dg \right| \leq \int_a^b |f| dg.$

Infine si può dimostrare la seguente

Proposizione 3.7 - *Sia $f \in C^0([a,b])$ e $g \in C^1([a,b])$, g monotona. Vale allora la seguente uguaglianza:*

$$\int_a^b f \, dg = \int_a^b f(x) g'(x) dx \ .$$

Si vede dunque che, nelle ipotesi enunciate, il simbolo dg si comporta esattamente come il differenziale di g. La proposizione 3.7 riconduce, in tal caso, il calcolo dell'integrale di Stieljes al calcolo di un ordinario integrale di Riemann.

Esercizi

1. Studiare la convergenza dei seguenti integrali impropri:

$$\int_0^{+\infty} \sin x \, dx \qquad \int_{-\pi/2}^{+\pi/2} \operatorname{tg} x \, dx \qquad \int_0^{\infty} \frac{(\sin x)^2}{x^2} dx$$

$$\int_0^{+\infty} \sin(x^2) dx \qquad \int_0^{+\infty} \cos(x^2) dx$$

(gli ultimi due sono detti integrali di Frésnel ed intervengono nella teoria della diffrazione della luce).

2. Dimostrare che, se $f : [a,b] \to \mathbf{R}$ è continua in (a,b) e integrabile in senso improprio su $[a,b]$, allora esiste $\xi \in (a,b)$ tale che

$$\int_a^b f(x) dx = (b-a) f(\xi) \ .$$

$f(\xi)$ viene detto *valor medio* di f in $[a,b]$.

Per esempio, verificare che il valor medio di $\frac{1}{\sqrt{x}}$ in $[0,1]$ è 2, il valor medio di $\frac{1}{\sqrt{1-x^2}}$ in $[-1,+1]$ è π.

3. Sia $f : [a,b] \to \mathbb{R}$ discontinua nei punti $\xi_1, \xi_2, \ldots, \xi_n$ ed esista l'integrale generalizzato $\int_a^b f(x)dx$. Posto

$$F(x) = \int_a^x f(t)dt = \int_a^{\xi_1} f(t)dt + \ldots + \int_a^{\xi_k} f(t)dt$$

se $\xi_k \leq x < \xi_{k+1}$ $(k = 1, 2, \ldots, n)$ dimostrare che:

a) $F(x)$ è continua in $[a,b]$.

b) $F(x)$ è derivabile nei punti di continuità di $f(x)$ e risulta $F'(x) = f(x)$.

c) Se $G(x)$ soddisfa le proprietà *a)* e *b)*, essa differisce da F per una costante.

4.* Se esiste $\displaystyle\int_a^{+\infty} f(x)dx$, è vero che $\displaystyle\lim_{x\to+\infty} f(x) = 0$?

5. Discutere, al variare del parametro reale α, la convergenza degli integrali

$$\int_a^b \frac{dx}{(a-x)|\log(a-x)|^\alpha}\ , \int_c^{+\infty} \frac{dx}{x(\log x)^\alpha}\ \ (c > 1)\ .$$

6. Trovare per quali valori del parametro reale x è convergente l'integrale

$$\Gamma(x) := \int_0^{+\infty} e^{-t} t^{x-1} dt\ .$$

Esso definisce una nuova funzione trascendente, detta *funzione gamma* di Eulero. Una delle proprietà più importanti della funzione Γ è (verificare!):

$$\Gamma(n+1) = n!\quad n \in \mathbb{N}\ .$$

Questa funzione perciò estende la definizione del fattoriale ai numeri reali (interpola il fattoriale).

7.* Ricavare, dal criterio di Cauchy per l'esistenza del limite finito per le funzioni (Proposizione 4.3.12) i seguenti criteri di convergenza per integrali impropri:

a) Sia $f : (a,b] \to \mathbb{R}$ integrabile su $[a+\delta, b]$ $\forall\ \delta > 0$; allora l'integrale improprio $\displaystyle\int_a^b f(x)dx$ esiste se e solo se, $\forall \varepsilon > 0$ si può trovare $\eta = \eta(\varepsilon)$ tale che:

$$\left|\int_{x'}^{x''} f(x)dx\right| < \varepsilon \qquad \forall\ x', x'' \in (a, a+\eta)\ .$$

b) Sia $f : [a, +\infty) \to \mathbb{R}$ integrabile su ogni intervallo limitato $[a, \omega]$ con $\omega \geq a$; allora l'integrale improprio $\displaystyle\int_a^{+\infty} f(x)dx$ esiste se e solo se, $\forall \varepsilon > 0$, si può trovare $\sigma = \sigma(\varepsilon)$

tale che:

$$\left|\int_{x'}^{x''} f(x)dx\right| < \varepsilon \quad \forall\, x', x'' \in (\sigma, +\infty)\ .$$

8*. *L'integrale in valor principale secondo Cauchy.*

Siano $I = [a,b]$ e $c \in (a,b)$. Sia poi $f : [a,c) \cup (c,b] \to \mathbb{R}$ una funzione integrabile sugli intervalli $[a, c-\delta]$ e $[c+\delta, b]$ per ogni $\delta > 0$. Come abbiamo detto, si dirà che f è integrabile (in senso improprio) in $[a,b]$ se esistono finiti i limiti

$$\lim_{\delta_1 \to 0_+} \int_a^{c-\delta_1} f \quad \text{e} \quad \lim_{\delta_2 \to 0_+} \int_{c+\delta_2}^{b} f\ .$$

ed in tal caso si pone

$$\int_a^b f = \lim_{\delta_1 \to 0_+} \int_a^{c-\delta_1} f + \lim_{\delta_2 \to 0_+} \int_{c+\delta_2}^{b} f\ .$$

Si noti che δ_1 e δ_2 tendono a zero indipendentemente.

Un caso particolare, che trova notevoli applicazioni in molti settori dell'Analisi matematica e della teoria delle equazioni differenziali, si ha assumendo $\delta_1 = \delta_2 = \delta$ e considerando il limite

$$\lim_{\delta \to 0_+} \left(\int_a^{c-\delta} f + \int_{c+\delta}^{b} f \right)\ .$$

Quando questo limite esiste finito esso si chiama *valor principale* (secondo Cauchy) dell'integrale di f su $[a,b]$ e viene indicato col simbolo

$$p.v. \int_a^b f \quad \text{o} \quad p.v. \int_a^b f(x)dx\ .$$

Chiaramente, se f è integrabile (in senso ordinario o in senso improprio) su $[a,b]$, l'integrale di f coincide con il suo valor principale.

Viceversa, il valor principale può esistere finito anche in casi in cui l'integrale improprio non esiste, come mostra il seguente esempio.

***Esempio* 3.12** - Siano $f(x) = \dfrac{1}{x}$ e $[a,b] = [-1,2]$.

Come abbiamo visto, f non è integrabile in senso improprio in un intorno dell'origine. Tuttavia, se $\delta > 0$, si ha:

$$\int_{-1}^{\delta} \frac{1}{x}dx + \int_{\delta}^{2} \frac{1}{x}dx = \int_{-1}^{\delta} \frac{1}{x}dx + \int_{\delta}^{1} \frac{1}{x}dx + \int_{1}^{2} \frac{1}{x}dx = \log 2$$

essendo $\dfrac{1}{x}$ dispari. Dunque $p.v. \int_{-1}^{2} \dfrac{1}{x}dx = \log 2$.

Appendice
CENNO ALL'ANALISI NON STANDARD

La prefazione alla prima edizione del libro "Non–standard Analysis" dell'americano Abraham Robinson inizia così:

"Nell'autunno del 1960 mi resi conto che i concetti e i metodi della Logica Matematica contemporanea potevano fornire un'intelaiatura adatta allo sviluppo del Calcolo Differenziale ed Integrale per mezzo di numeri infinitamente piccoli ed infinitamente grandi...

... La teoria risultante fu chiamata da me Analisi non–standard perché coinvolge ed era, in parte, ispirata dai cosiddetti modelli non–standard dell'Aritmetica la cui esistenza era stata sottolineata per primo da T. Skolem. ..."

Ovviamente non possiamo descrivere i risultati di logica ai quali Robinson fa riferimento.

Una domanda sorge spontanea: che ragione c'è di introdurre l'Analisi non–standard? È forse "più potente" dell'Analisi moderna? (chiamiamo così, per intenderci, quella da Weierstrass (1870) in poi).

È stato dimostrato che le due teorie sono equivalenti nel senso che *ogni* teorema di Analisi non–standard ha un equivalente in Analisi moderna e viceversa. La ragione è diversa e, forse, è racchiusa in alcune frasi dell'eminente logico Kurt Gödel, inserite nella prefazione alla seconda edizione del libro di Robinson (1974), che qui riportiamo parzialmente:

"...l'Analisi non–standard spesso semplifica in modo sostanziale le dimostrazioni, non solo di teoremi elementari, ma anche di risultati profondi. Questo è vero, ad esempio, ...

... Questi dati di fatto dovrebbero impedire una interpretazione sbagliata, piuttosto comune, dell'analisi non–standard, vale a dire che essa sia un tipo di stravaganza o di moda passeggera dei logici matematici. Niente potrebbe essere più lontano dalla verità. Piuttosto, ci sono buone ragioni per credere che l'analisi non–standard, in una versione o nell'altra, sarà l'analisi del futuro".

Il nostro scopo è di illustrare brevemente le idee principali che hanno permesso di arricchire il campo dei reali con i numeri che Leibniz chiamava "ideali": gli infinitesimi e gli infiniti. L'allievo ricorderà che abbiamo già parlato di infinitesimi e di infiniti come di funzioni che tendono a zero o all'infinito, rispettivamente. Per Leibniz e Robinson, come vedremo, essi sono invece *numeri* veri e propri; in essi non c'è nulla di "dinamico".

Ci si può chiedere a questo punto, quale sia la ragione per "reinventare" numeri infinitesimi ed infiniti dal momento che Leibniz e tutti gli studiosi del Calcolo differenziale usavano già tali entità ideali.

La ragione è molto semplice: non si conosceva un modo rigoroso per introdurre quei numeri ed operare con essi; in ultima analisi, nonostante i molti successi della teoria Leibniziana (si pensi alla meccanica dei corpi celesti), questo fatto

significava una intrinseca limitazione alle possibilità di crescita ed espansione della teoria stessa.

Può essere interessante vedere come Eulero, nella sua "Opera Omnia" vol. primo, 1748, derivi la formula di Mac Laurin per il coseno: sia n intero; dalla formula di De Moivre e da quella del binomio di Newton si ricava:

$$\begin{aligned}\cos nx &= \frac{1}{2}\{(\cos x + i\sin x)^n + (\cos x - i\sin x)^n\} = \\ &= \frac{1}{2}\{(\cos x)^n + ni\sin x(\cos x)^{n-1} + \frac{n(n-1)}{2}(i\sin x)^2(\cos x)^{n-2} + \\ &+ \ldots + (i\sin x)^n + [(\cos x)^n - ni\sin x(\cos x)^{n-1} + \\ &+ \frac{n(n-1)}{2}(-i\sin x)^2(\cos x)^{n-2} + \ldots + (-i\sin x)^n]\} = \\ &= \frac{1}{2}\{2(\cos x)^n - n(n-1)(\sin x)^2(\cos x)^{n-2} + \ldots\} .\end{aligned}$$

Sia ora x "infinitamente piccolo"; sarà $\cos x = 1, \sin x = x$; sia poi n infinitamente grande (cosicché $n - 1 = n$) in modo che nx sia di grandezza finita; posto $nx = v$ si ricava:

$$\cos v = 1 - \frac{1}{2}v^2 + \ldots .$$

A quel tempo non c'era modo di giustificare tutti i passaggi, in particolare $\cos x = 1$, $\sin x = x$ per x infinitamente piccolo. Sotto l'incalzare delle numerose critiche, gradualmente la teoria degli infinitesimi di Leibniz fu abbandonata per lasciare posto all'analisi moderna, per oltre un secolo.

Arriviamo dunque a Robinson. Un aspetto estremamente interessante emerso dalle scoperte di logica è che sono possibili moltissime estensioni non–standard dell'analisi moderna: Robinson elabora una di esse, che ha l'importante proprietà di essere collegata all'analisi moderna in modo tale da poter interpretare e trasferire ogni tipo di definizione o risultato dall'una all'altra. Non possiamo certo presentare la costruzione originale di Robinson; ci limitiamo alla costruzione di un modello non–standard dei numeri reali, che indicheremo con ${}^*\mathbb{R}$ (da non confondere con $\mathbb{R}^*$!), nel modo più elementare possibile.

Accenneremo poi alle definizioni non–standard di continuità, di derivata ed integrale.

Costruzione di ${}^*\mathbb{R}$.

Per capire l'idea base ritorniamo alla costruzione di Cantor dei numeri reali, accennata nel paragrafo 4.4.3. In questa costruzione si considerano successioni fondamentali di numeri razionali, $\{r_n\} \subset \mathbb{Q}$, quelle cioè che soddisfano la condizione di Cauchy.

Il punto chiave della costruzione di $\mathbb{R}$ sta nell'assegnare una relazione di equivalenza nell'insieme delle successioni: due di esse, $\{r_n\}$ e $\{s_n\}$, sono ritenute equivalenti se $r_n - s_n \to 0$ per $n \to +\infty$.

Successioni equivalenti individuano lo *stesso numero* reale. Come conseguenza, le successioni seguenti, equivalenti nel senso precisato sopra, individuano lo stesso numero reale: lo zero:

$$\{0\},\ \{2^{-n}\},\ \left\{\frac{1}{n}\right\},\ \left\{\frac{1}{n^2}\right\},\ldots,\left\{\frac{1}{n^k}\right\},\ldots\ . \tag{A.1}$$

D'altra parte, queste successioni risultano molto diverse tra loro per quanto riguarda la velocità con la quale tendono a zero. La relazione di equivalenza introdotta da Cantor non segnala tale differenza.

L'idea di Robinson è di partire da successioni di reali e di introdurre nel loro insieme una relazione di equivalenza, in modo tale da tener conto di tale diversità.

Si potrebbe pensare di considerare come "numeri" le singole successioni oppure di ritenerle equivalenti se differiscono solo per i valori assunti su un insieme finito di interi. Questo modo di agire è giustificato dal fatto che, dopotutto siamo interessati al comportamento delle successioni per $n \to +\infty$.

Per le classi di equivalenza così ottenute si definiscono somma e prodotto nella solita maniera:

$$[r_n] + [s_n] := [r_n + s_n]$$

$$[r_n] \cdot [s_n] := [r_n \cdot s_n]\ .$$

Ci si accorge però subito che non possono valere le ordinarie regole di calcolo. Infatti, ad esempio, $[0]$ è l'elemento neutro della somma e le classi individuate dalle successioni

$$r_n = 0, 1, 0, 3, 0, 5, \ldots$$

$$s_n = 0, 0, 2, 0, 4, 0, \ldots$$

sono diverse da $[0]$. Tuttavia è evidente che $[r_n] \cdot [s_n] = [0]$. Non vale quindi la legge di annullamento del prodotto.

Occorre una relazione di equivalenza più "bilanciata" nel senso che non deve identificare "troppe" successioni (come quella di Cantor) altrimenti non si esce dal campo dei reali e al contempo non "troppo poche" altrimenti otterremmo un insieme sicuramente più grande di $\mathbb{R}$ ma nel quale non possiamo operare con le ordinarie regole di calcolo.

Il primo passo è di operare una suddivisione dei sottoinsiemi di $\mathbb{N}$ in "grandi" e "piccoli".

A tale scopo introduciamo una funzione m definita sui sottoinsiemi di $\mathbb{N}$ con le seguenti proprietà:

i) $\forall A \subseteq \mathbb{N}$, $m(A) = 0$ oppure $m(A) = 1$;

ii) $m(\emptyset) = 0$, $m(\mathbb{N}) = 1$;

iii) se $A \subset \mathbb{N}$ è finito allora $m(A) = 0$;

iv) se $A_1, A_2, \ldots, A_k$ sono sottoinsiemi di $\mathbb{N}$ a due a due disgiunti allora

$$m\left(\bigcup_{j=i}^{k} A_j\right) = \sum_{j=i}^{k} m(A_j) \quad \text{(proprietà di additività)} .$$

Una funzione come m si chiama *misura finitamente additiva* (per la *iv)*) su $\mathcal{P}(\mathbb{N})$.

L'esistenza di m (anzi di infinite m) è una conseguenza dell'assioma di scelta (cfr. appendice al Cap. 1).

Definizione A.1 - *Sia $A \subseteq \mathbb{N}$. Diciamo che A è piccolo se $m(A) = 0$, che A è grande se $m(A) = 1$.*

Sia ora $\mathbb{R}^{\mathbb{N}}$ l'insieme delle successioni a valori reali.

In $\mathbb{R}^{\mathbb{N}}$ introduciamo la seguente relazione di equivalenza che indichiamo con il simbolo $\overset{*}{\sim}$:

$$\{r_n\} \overset{*}{\sim} \{s_n\} \text{ se e solo se } m\{i \in \mathbb{N} : r_i = s_i\} = 1$$

ovvero se l'insieme degli indici su cui coincidono è un sottoinsieme *grande* di $\mathbb{N}$.

Definizione A.2 - ${}^*\mathbb{R} := \mathbb{R}^{\mathbb{N}} / \overset{*}{\sim}$.

L'insieme ${}^*\mathbb{R}$ si chiama a volte insieme degli *iperreali*. I suoi elementi sono classi di equivalenza (rispetto a $\overset{*}{\sim}$) di successioni a valori reali, che indicheremo con il simbolo $< \cdot >$. Useremo lettere greche per indicare iperreali; così, $\alpha \in {}^*\mathbb{R}$ significa che $\alpha = < r_n >$ per qualche successione $r_n : \mathbb{N} \to \mathbb{R}$. Introduciamo una relazione d'ordine in ${}^*\mathbb{R}$, che continueremo ad indicare con il simbolo $<$:

se $\alpha = < r_n >, \beta = < s_n >$ allora $\alpha < \beta$ se e solo se $m\{i \in \mathbb{N} : r_i < s_i\} = 1$.

Le operazioni di somma e prodotto (indicate ancora con $+$ e $\cdot$) si definiscono ponendo

$$\alpha + \beta := < r_n + s_n > \quad \text{e} \quad \alpha \cdot \beta := < r_n \cdot s_n > \; .$$

Proposizione A.1 - *${}^*\mathbb{R}$ è un campo ordinato.*

Come conseguenza, in ${}^*\mathbb{R}$ valgono tutte le usuali regole di calcolo. In ${}^*\mathbb{R}$, $\mathbb{R}$ è immerso in modo naturale, nel senso che ${}^*\mathbb{R}$ contiene un campo ordinato *completo*, che per il teorema di isomorfismo, (vedi 2.2.5), possiamo senz'altro identificare con $\mathbb{R}$.

È questo l'insieme degli elementi di ${}^*\mathbb{R}$ della forma ${}^*x := < x >$, classe di equivalenza individuata dalla successione costante $r_n = x, \forall n \in \mathbb{N}$. Scriveremo senz'altro $\mathbb{R} \subset {}^*\mathbb{R}$ e x al posto ${}^*x$.

La proprietà che più ci interessa è però che ${}^*\mathbb{R}$ è *una estensione propria* di $\mathbb{R}$, cioè $\mathbb{R} \subset {}^*\mathbb{R}$ (un *allargamento* nella terminologia di Robinson).

Siano infatti $\alpha = < \frac{1}{n} >$ ed $x \in \mathbb{R}_+$. Allora $\alpha > 0$ e l'insieme

$$\{i \in \mathbb{N} : \frac{1}{i} \geq x\}$$

è vuoto oppure costituito da un insieme finito di indici e pertanto è *piccolo* nel senso della definizione A.1.

Ne segue che $0 < \alpha < x$; poiché x è arbitrario in $\mathbb{R}_+$ si ha: $0 < \alpha < x$, $\forall x \in \mathbb{R}_+$. Ciò implica $\alpha \notin \mathbb{R}$; numeri come α si chiamano infinitesimi. Più precisamente:

Definizione A.2 - *Si dice che* $\varepsilon \in {}^*\mathbb{R}$ *è infinitesimo se* $|\varepsilon| < x, \forall x \in \mathbb{R}_+$. *Si dice che* $\omega \in {}^*\mathbb{R}$ *è infinito se* $|\omega| > x \ \forall \ x \in \mathbb{R}_+$.

Un esempio di infinito è $< n >$. Notiamo che $\beta = < \frac{1}{n^2} >$ è infinitesimo e risulta $\beta < \alpha$, dove α è l'infinitesimo $< \frac{1}{n} >$, come facilmente si verifica.

Così pure $< n^2 >$ è un infinito maggiore di $< n >$.

Per indicare che ε è infinitesimo scriveremo $\varepsilon \approx 0$; l'unico infinitesimo reale è lo zero. Due numeri α e β si dicono infinitamente vicini, e si scrive $\alpha \approx \beta$, se $\alpha - \beta \approx 0$. È poi facile verificare che:

$$\varepsilon \approx 0, \eta \approx 0, x \in \mathbb{R} \ \Rightarrow \ \varepsilon + \eta \approx 0, \ \varepsilon \cdot \eta \approx 0, \ \varepsilon \cdot x \approx 0 \ ;$$

$$\varepsilon \approx 0, \varepsilon \neq 0 \ \Rightarrow \ \frac{1}{\varepsilon} \text{ infinito } ;$$

$$\omega, \nu \text{ infiniti positivi}, x \in \mathbb{R}, x \neq 0 \ \Rightarrow \ \omega + \nu, \ \omega \cdot x \text{ infiniti } ;$$

$$\omega \text{ infinito } \Rightarrow \frac{1}{\omega} \approx 0.$$

È chiaro che ${}^*\mathbb{R}$ *non* può essere completo, altrimenti sarebbe isomorfo ad $\mathbb{R}$; infatti l'insieme degli infinitesimi non può avere estremo superiore (e inferiore) in ${}^*\mathbb{R}$, come facilmente si verifica (*). Né ${}^*\mathbb{R}$ può soddisfare la proprietà di Archimede: infatti se $x \in \mathbb{R}_+$ ed $\varepsilon \approx 0, \varepsilon > 0$, non esiste $n \in \mathbb{N}$ tale che $n\varepsilon > x$.

I numeri di ${}^*\mathbb{R}$ si possono suddividere in finiti ed infiniti.

Definizione A.3 - $\alpha \in {}^*\mathbb{R}$ *è finito se* $\exists r \in \mathbb{R}_+$ *tale che* $|\alpha| < r$.

Ora, tutti i numeri finiti sono della forma $x + \varepsilon$ con $x \in \mathbb{R}$ ed $\varepsilon \approx 0$; più precisamente vale il seguente teorema, fondamentale per lo sviluppo del calcolo differenziale ed integrale.

■ **Teorema A.2** - *Sia* $\alpha \in {}^*\mathbb{R}$, α *finito. Allora esiste uno ed un solo numero reale* x *tale che* $x - \alpha \approx 0$; x *si chiama parte standard di* α *e si indica con* $st(\alpha)$ *oppure* ${}^\circ\alpha$.

(*) Sia E l'insieme degli infinitesimi. Per assurdo supponiamo che esista $\alpha \in {}^*\mathbb{R}$ tale che $\alpha = \sup E$. Si hanno due possibilità: $\alpha \approx 0$ oppure $\alpha \not\approx 0$. Se $\alpha \approx 0$, per ogni altro infinitesimo positivo ε si ha $\alpha + \varepsilon \approx 0$ e $\alpha + \varepsilon > \alpha$: contraddizione. Se $\alpha \not\approx 0$ ed ε è un infinitesimo positivo si ha $\alpha - \varepsilon > \eta \ \forall \eta \in E$: ancora contraddizione.

Se $x \in \mathbb{R}$, l'insieme dei punti della forma $x + \varepsilon$ con $\varepsilon \approx 0$ si chiama con il termine leibniziano *monade* (o anche alone) di x e si indica con $\mu(x)$.

Una visione suggestiva della retta iperreale ${}^*\mathbb{R}$, potrebbe essere la seguente.

Fig. 8.8

Ritorniamo ora al procedimento di estensione da $\mathbb{R}$ a ${}^*\mathbb{R}$. Esso si può applicare a qualunque sottoinsieme di $\mathbb{R}$, estendendolo a un sottoinsieme di ${}^*\mathbb{R}$: se $E \subset \mathbb{R}$, ${}^*E$ è definito dagli elementi $\alpha \in {}^*\mathbb{R}$ tali che, se $\alpha = \langle r_n \rangle$ allora

$$m\{i \in \mathbb{N}: r_i \in E\} = 1 \ .$$

Così all'intervallo $[a, b]$ corrisponde l'intervallo ${}^*[a, b]$ che contiene la monade destra di a, quella sinistra di b e le monadi di tutti gli elementi di ${}^*(a, b)$.

$\mathbb{N}, \mathbb{Z}, \mathbb{Q}$ si estendono a ${}^*\mathbb{N}, {}^*\mathbb{Z}, {}^*\mathbb{Q}$. In particolare ${}^*\mathbb{N}$ è costituito dagli interi di $\mathbb{N}$ e dagli interi infiniti: questi ultimi sono rappresentati da successioni del tipo $< n >, < n^2 >, \ldots$.

I sottoinsiemi di ${}^*\mathbb{R}$ della forma ${}^*E$ con $E \subseteq \mathbb{R}$ si chiamano *insiemi standard*. Certamente essi non sono i soli sottoinsiemi di ${}^*\mathbb{R}$; ad esempio l'insieme degli infinitesimi non è standard; non possiamo però addentrarci più in profondità in questa breve introduzione.

Tramite la nozione di parte standard si possono introdurre le analoghe non standard delle nozioni topologiche del capitolo terzo.

Ad esempio, sia $A \subseteq {}^*\mathbb{R}$, A standard.

A è aperto se per ogni $\alpha \in A, \alpha$ finito $\Rightarrow \mu(\alpha) \subset A$.

A è chiuso se per ogni $\alpha \in A, \alpha$ finito $\Rightarrow st(\alpha) \in A$.

A è compatto se ogni $\alpha \in A$ è finito e $st(\alpha) \in A$.

Gli enti fondamentali dell'analisi sono le funzioni ed anch'esse si possono estendere con la procedura "$*$".

Infatti, $f : E \to \mathbb{R}$, $E \subseteq \mathbb{R}$, si può estendere in modo naturale a ${}^*f : {}^*E \to {}^*\mathbb{R}$. Ad esempio $f(x) = x^2$ si estende a ${}^*f(\alpha) = \alpha^2$,$f(x) = \exp(x)$ si estende a $f^*(\alpha) = \exp(\alpha)$; come si vede la "corrispondenza" opera nello stesso modo anche sugli elementi non reali di ${}^*A$. Le funzioni ottenute in questo modo si chiamano *standard* e scriveremo senz'altro f invece di ${}^*f$.

Veniamo ora alle definizioni di continuità, derivata ed integrale per una funzione standard $f : {}^*(a, b) \to {}^*\mathbb{R}$.

Continuità : sia $x \in {}^*(a, b) \cap \mathbb{R}$; f si dice continua in x se

$$f(x + \varepsilon) \approx f(x) \quad \forall \varepsilon \approx 0$$

ovvero se $st[f(x + \varepsilon)] = f(x), \quad \forall \varepsilon \approx 0.$

Continuità uniforme: f è continua uniformemente in $^*(a,b)$ *se* $x', x'' \in^*(a,b)$ *e* $x' \approx x'' \Rightarrow f(x') \approx f(x'')$.

Derivata. Sia $x \in^*(a,b) \cap \mathbb{R}$; *il numero reale* $c \in \mathbb{R}$ *si dice derivata di f in x e si scrive* $c = f'(x)$ *se*

$$\frac{f(x+\varepsilon)-f(x)}{\varepsilon} \approx c \quad \forall \varepsilon \approx 0 ,$$

ovvero se $st\left[\frac{f(x+\varepsilon)-f(x)}{\varepsilon}\right] = c \ , \quad \forall \varepsilon \approx 0.$

Integrale. Ci limitiamo a funzioni continue in $^*[a,b]$. Sia $\omega \in^* \mathbb{N}\setminus\mathbb{N}$ un intero infinito e poniamo $\alpha_k = a + \frac{k}{\omega}(b-a), k = 1, 2, \ldots, \omega$. I punti α_k costituiscono una suddivisione infinita di $[a,b]$, tale che $\alpha_k - \alpha_{k-1} = \frac{b-a}{\omega} \approx 0$. Poniamo

$$S_\omega(f) = \sum_{k=1}^{\omega} f(\alpha_k)(\alpha_k - \alpha_{k-1}) = \frac{b-a}{\omega} \sum_{k=1}^{\omega} f(\alpha_k) .$$

Si osservi che $S_\omega(f)$ è somma di infiniti termini infinitesimi.
Il numero reale I si dice integrale di f in $[a,b]$ se

$$S_\omega(f) \approx I \quad \forall \omega \in^*\mathbb{N}\setminus\mathbb{N} ,$$

ovvero se $st[S_\omega(f)] = I \ , \quad \forall \omega \in^* \mathbb{N}\setminus\mathbb{N}$.
Come si vede l'operazione di parte standard sostituisce l'operazione di limite che qui non è necessaria.

Terminiamo questo flash sull'Analisi non–standard con una dimostrazione allo scopo di dare un'idea, seppure minima, di ciò che Gödel intendeva con le parole riportate all'inizio. È la dimostrazione del teorema di Cantor–Heine: *ogni funzione continua su un compatto* è *uniformemente continua.*

Dimostrazione (non standard). Sia $K \subset^* \mathbb{R}$ compatto e $f : K \to^* \mathbb{R}$ continua. Se x' e $x'' \in K$ e $x' \approx x''$ sia $x = st(x') = st(x'')$. Allora $x \in K$ essendo K compatto ed essendo f continua si ha:

$$f(x') \approx f(x) \approx f(x'') . \ \square$$

1. $f(x) = H(x)$: gradino di Heaviside

2. $f(x) = \mathrm{sgn}(x)$: segno di x

3. $f(x) = |x|$

4. $f(x) = x_+ = \max\{0, x\}$

5. $f(x) = [x]$

6. $f(x) = (x) = x - [x]$ = mantissa di x

13
f(x)= √(1 − x²)
14
exp(−x)
exp(x)
15
0.6 0.5 0.3̄ 0.1 10 3 2 1.6̄
0.8
1.25
1
f(x)= aˣ ,a reale >0
16
Ch x
½ eˣ
Sh x
17
f(x)= Th(x)
18
f(x)= logx

25

1/e

f(x)= xlogx

26

1/e

$f(x)= x^x = \exp(x\log x)$

27

f(x)= sinx

28

f(x)= cosx

29

la relazione: x → arcsin(x)

30

la relazione: x → arccos(x)

31
f(x) = tg(x)
32
la relazione: x → arctg(x)
33
f(x)= (sinx)^2
f(x)= |sinx|
34
f(x)= (sinx)^3
f(x)= sinx
f(x)= √sin x
f(x)= sinx
36
f(x) = sin x / x

37

$f(x) = \sin(x^2)$

38

$f(x) = x \sin x$

39

$f(x) = \sin(1/x)$

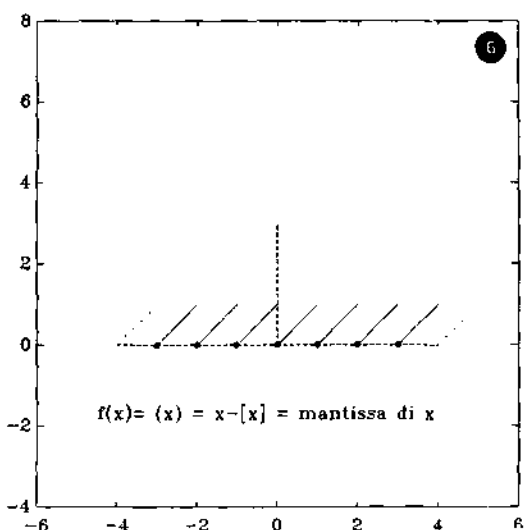
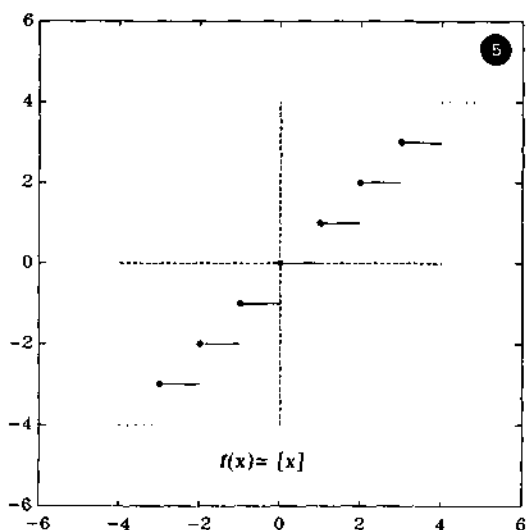
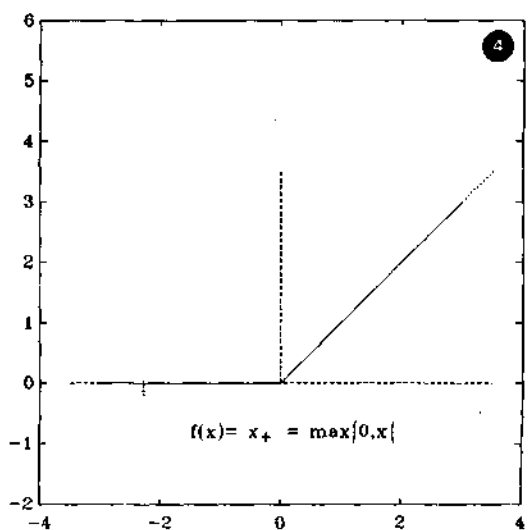
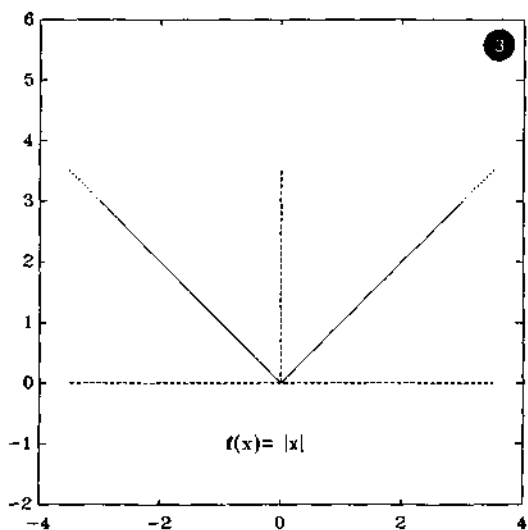
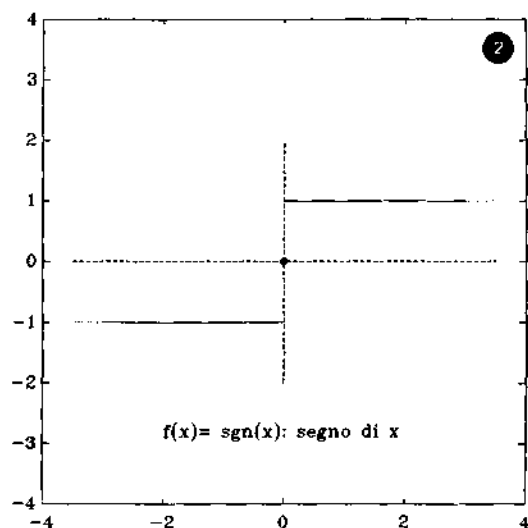
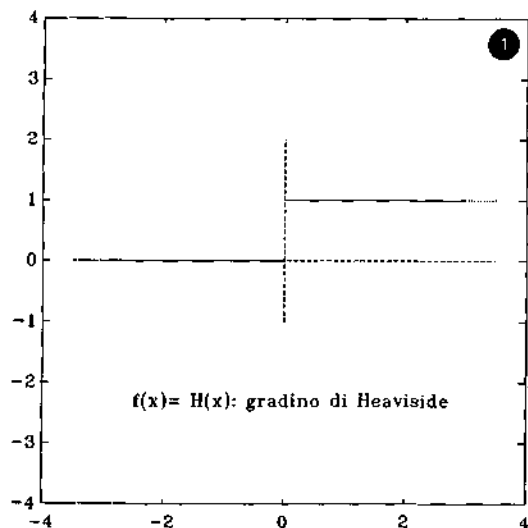
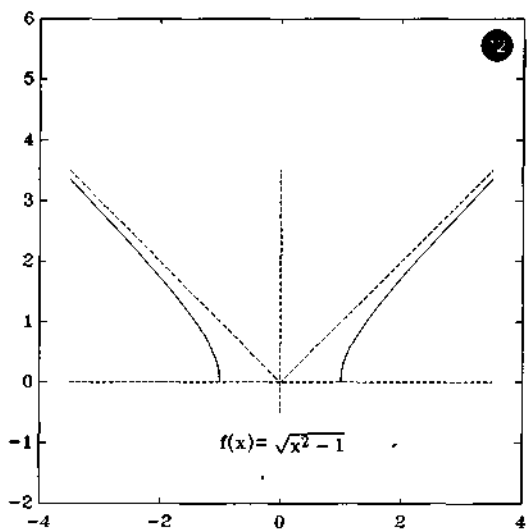
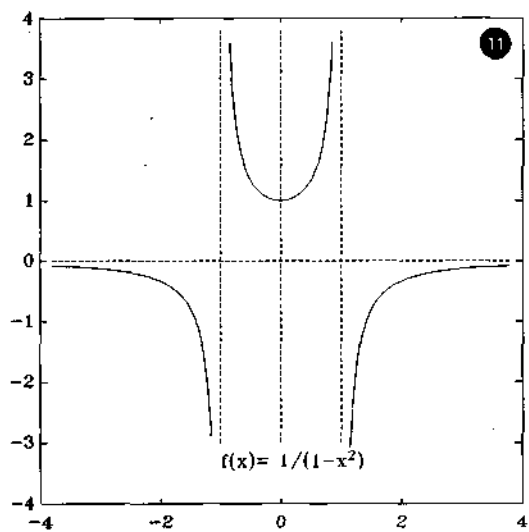
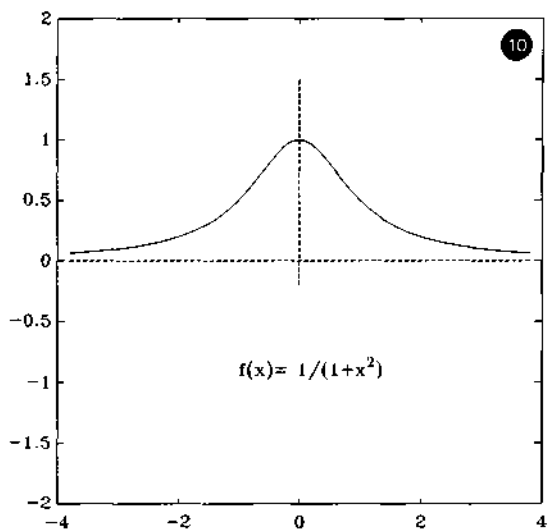
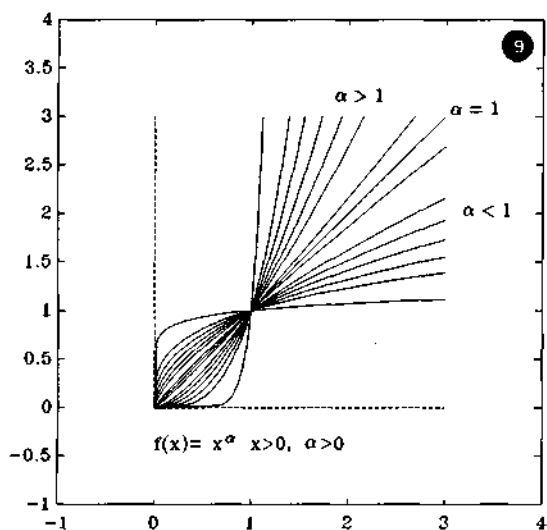
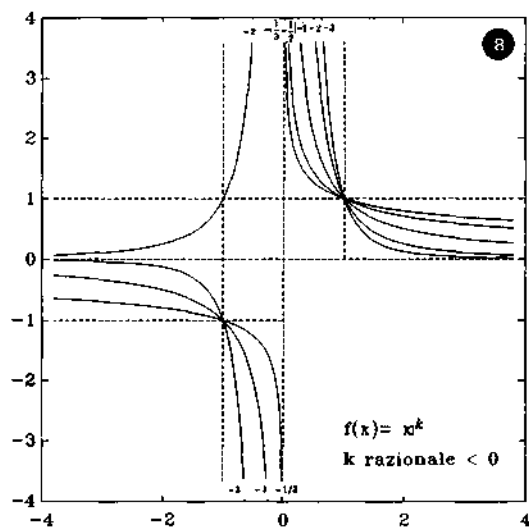
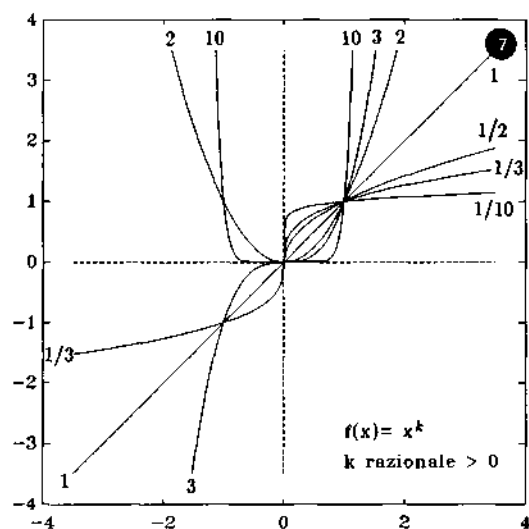
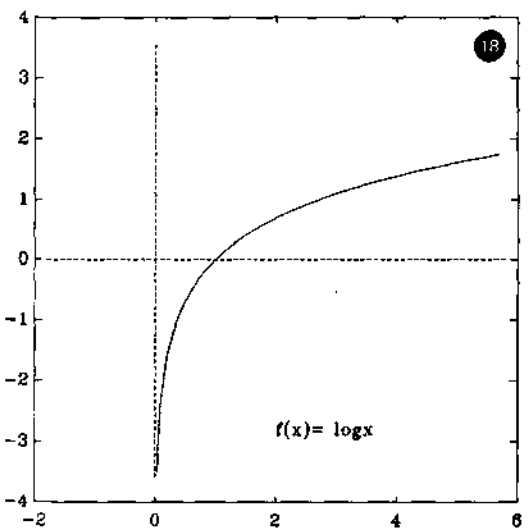
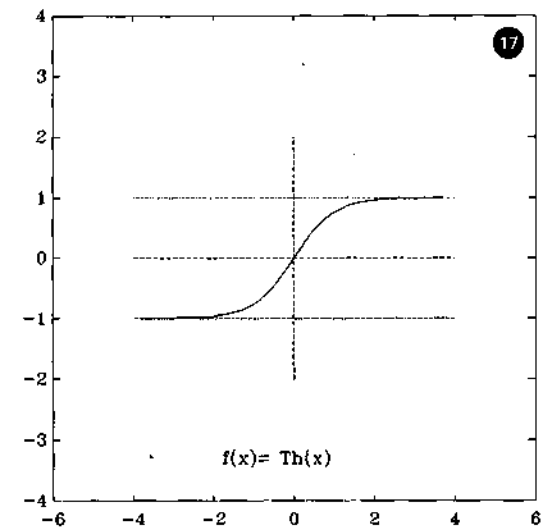
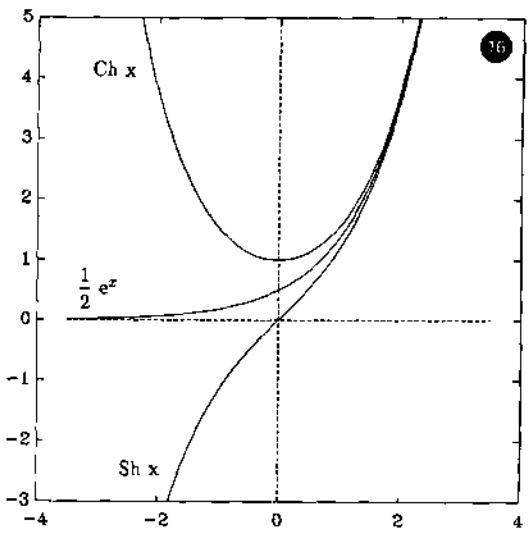
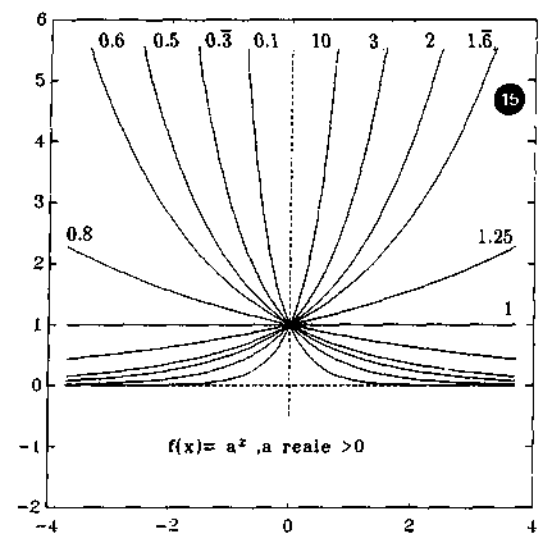
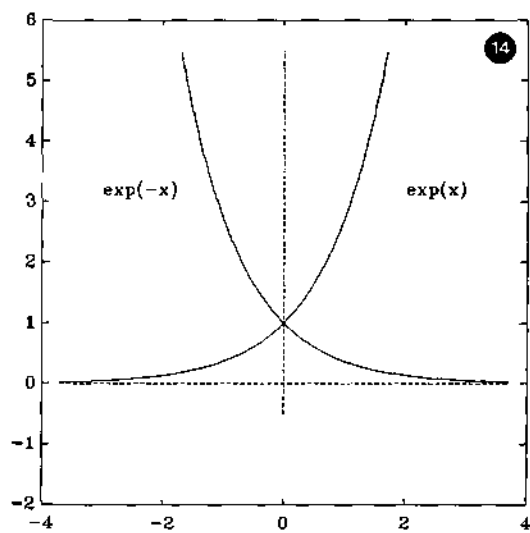
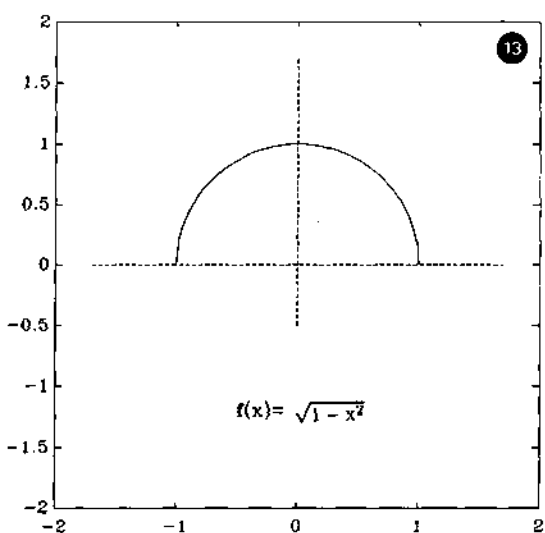
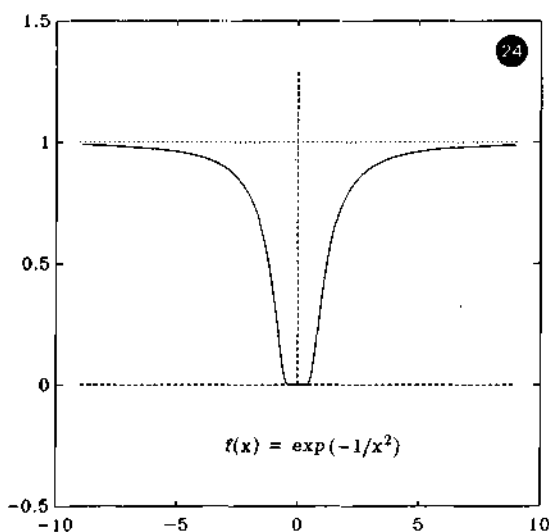
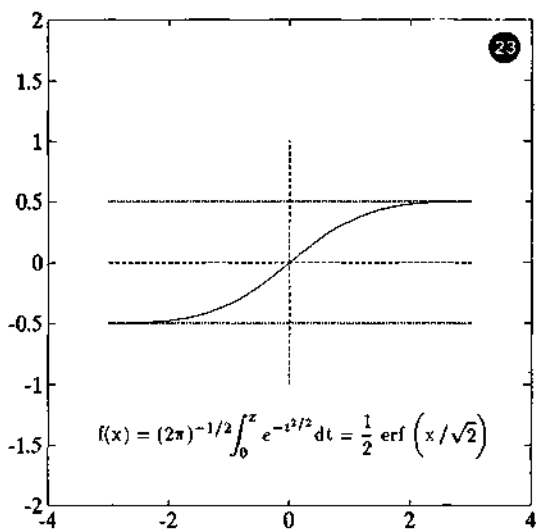
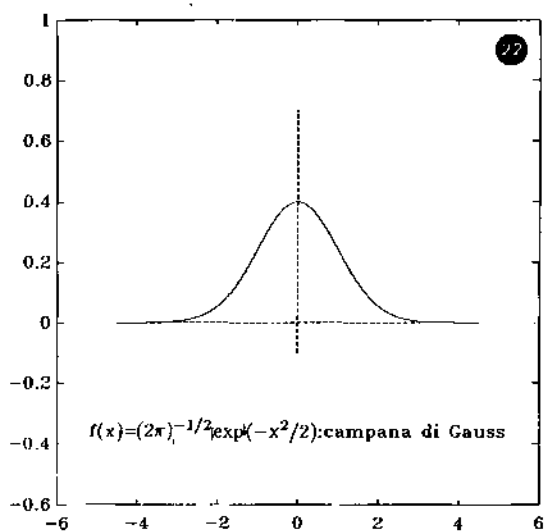
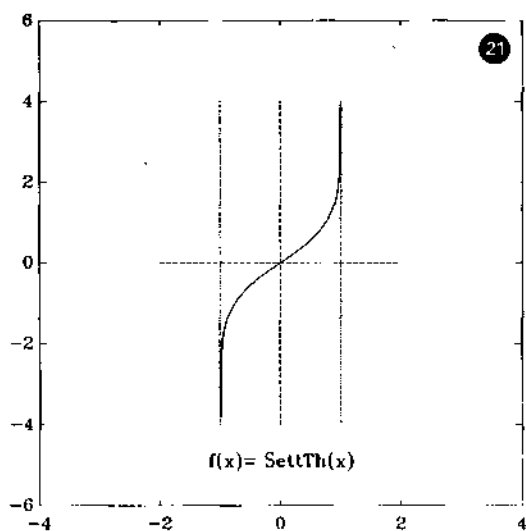
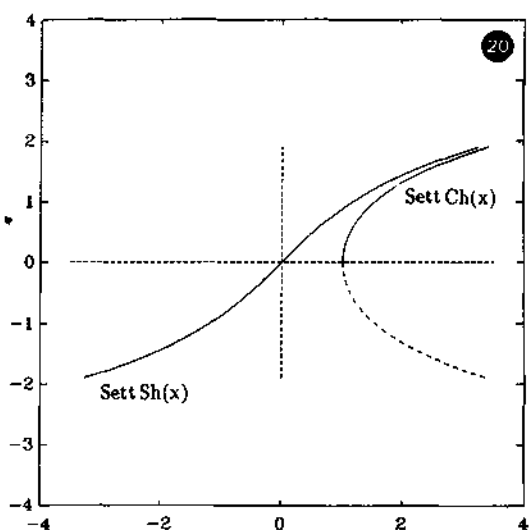
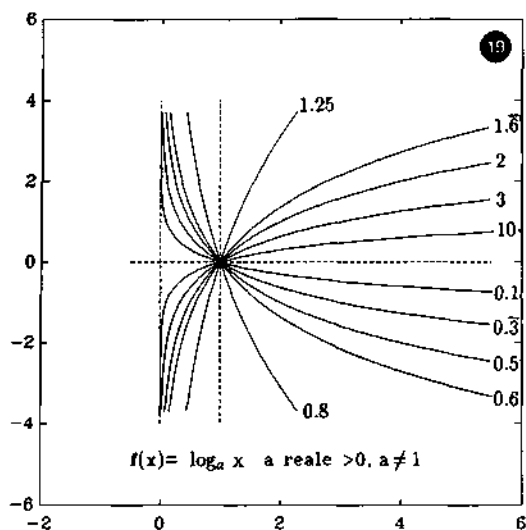
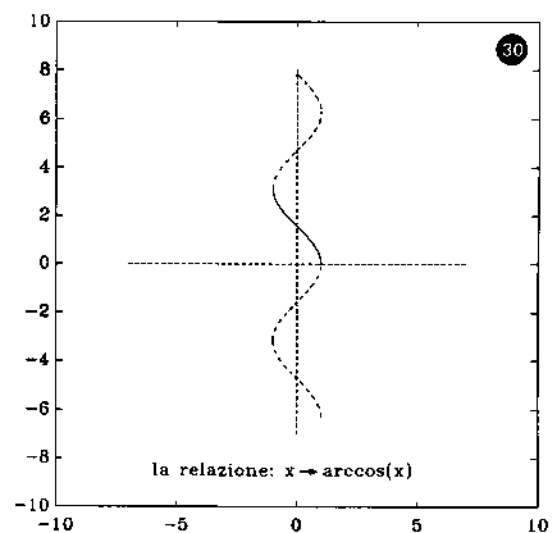
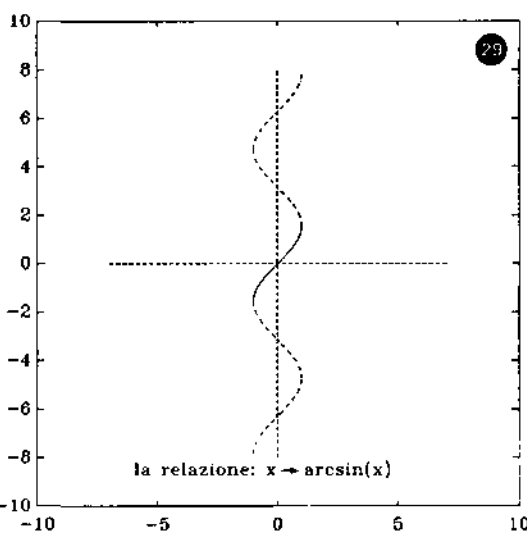
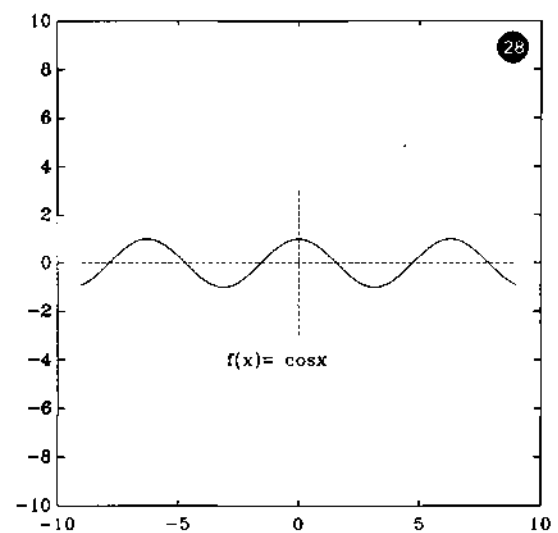
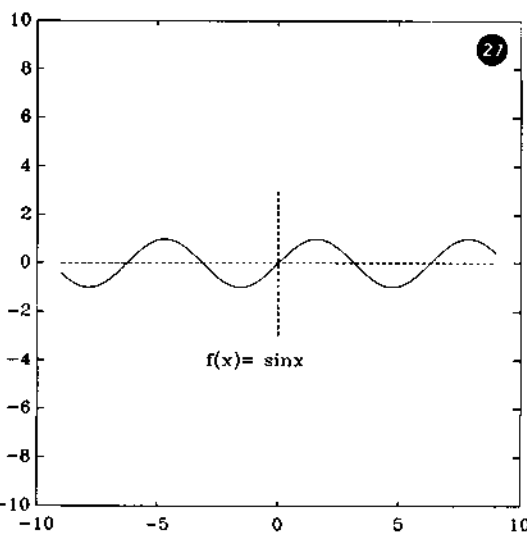
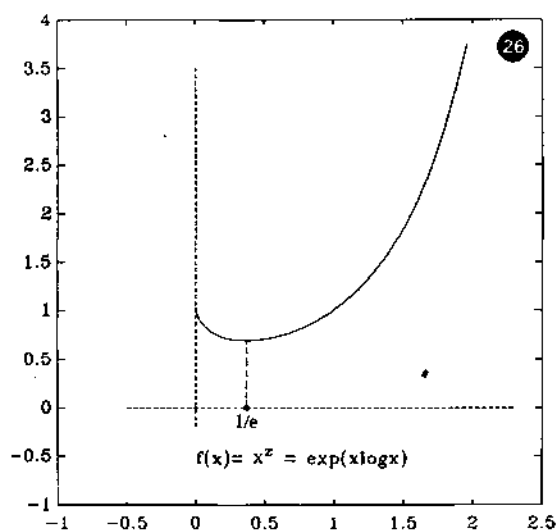
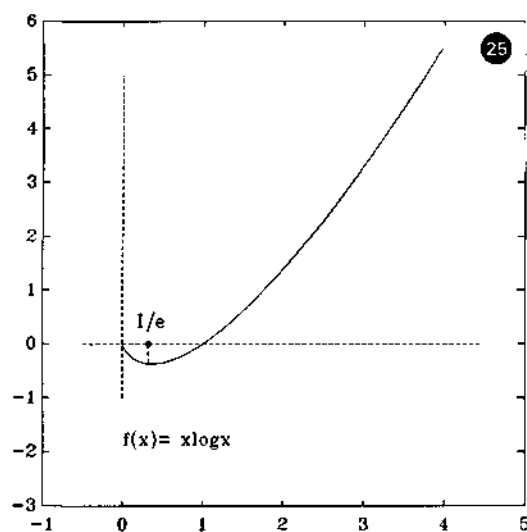
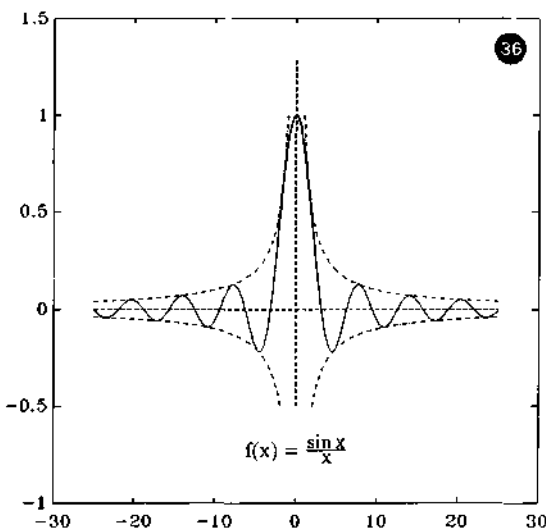
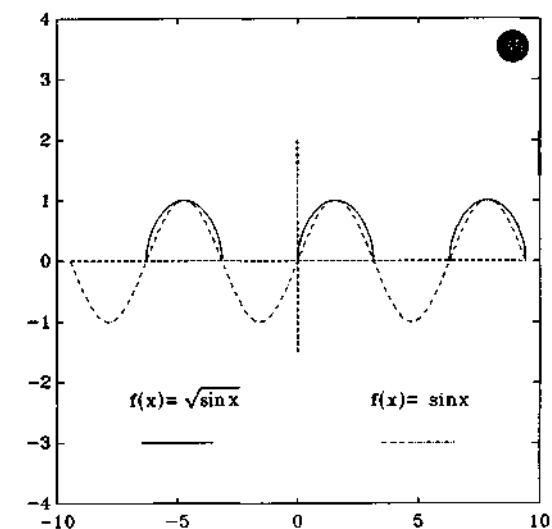
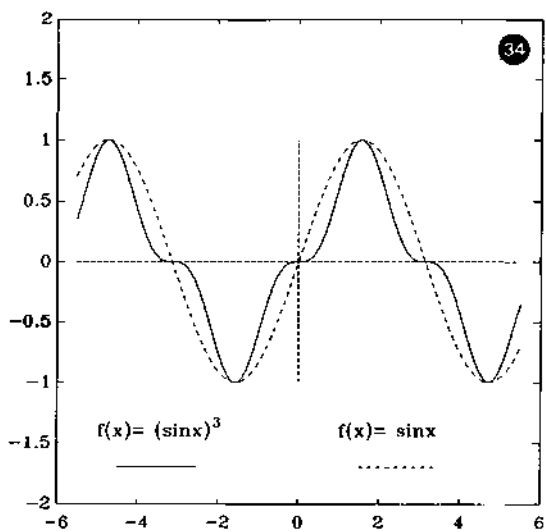
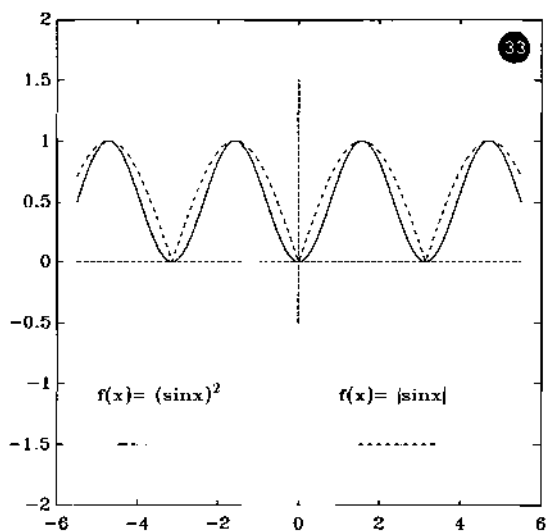
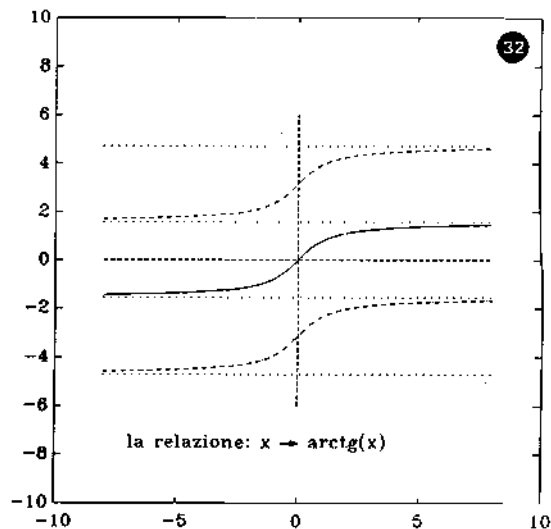
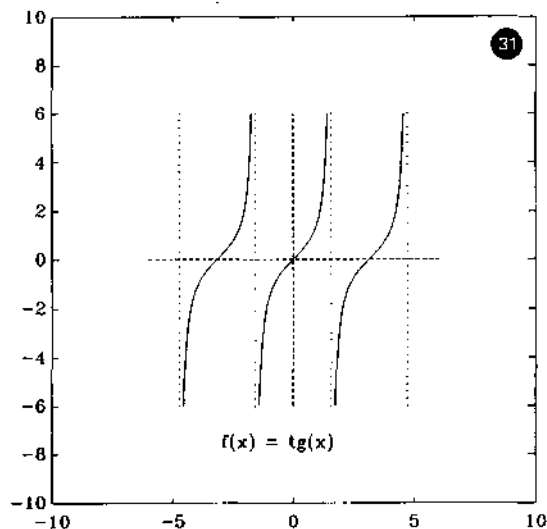
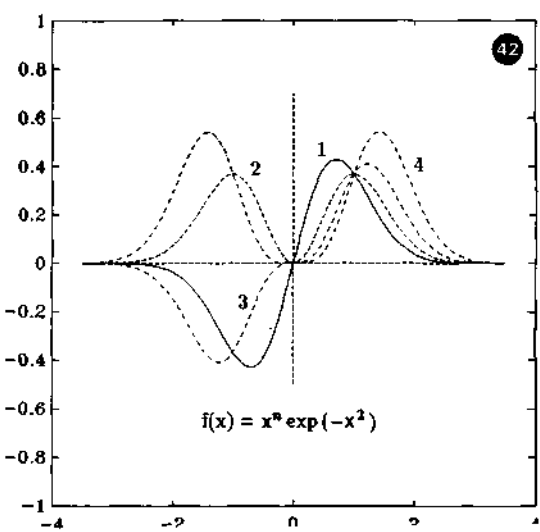
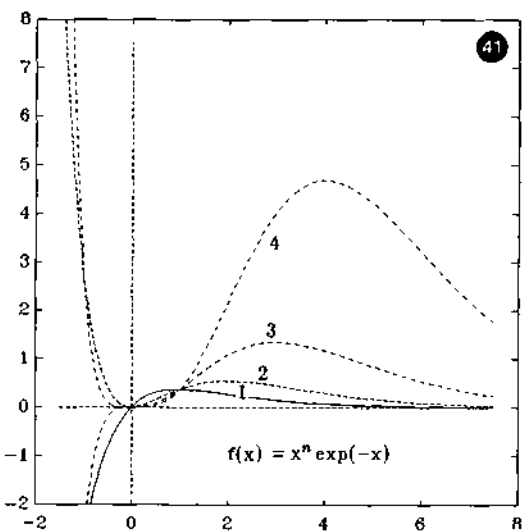
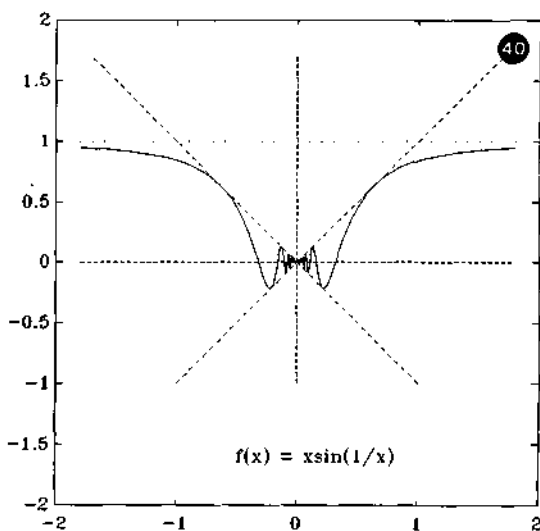
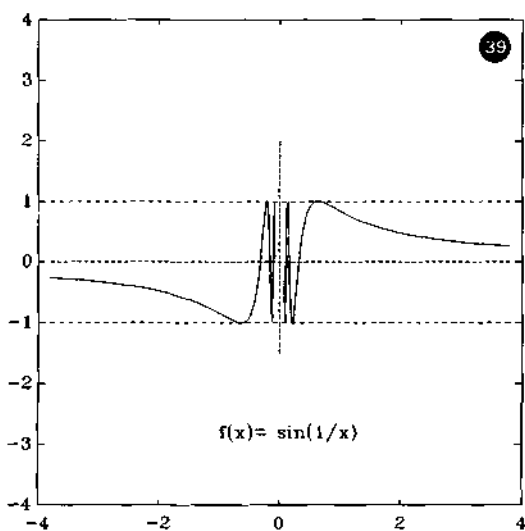
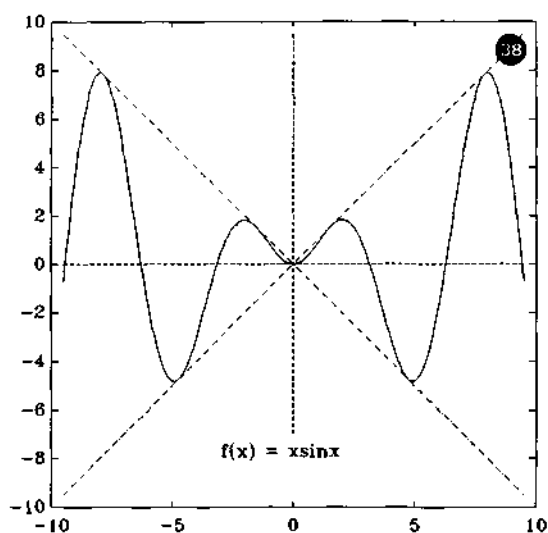
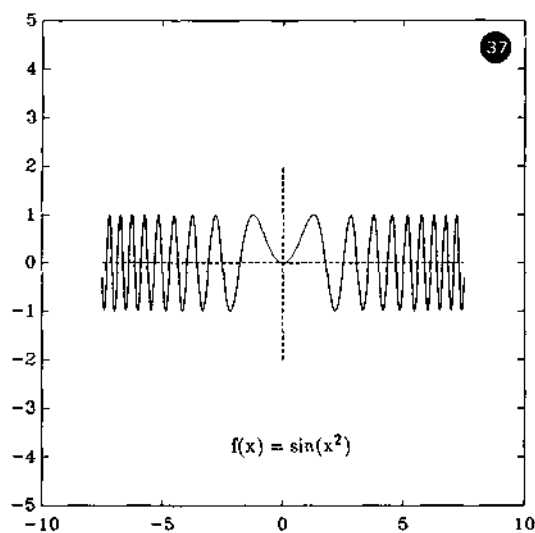
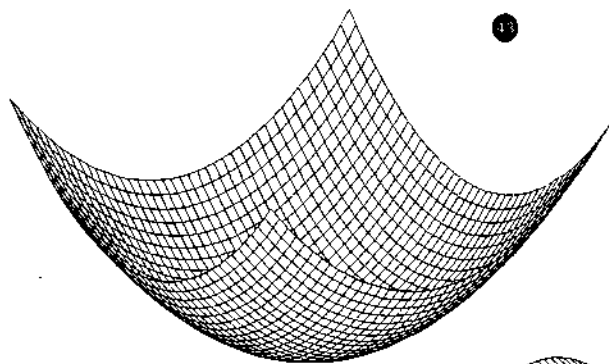
40

$f(x) = x \sin(1/x)$

41

$f(x) = x^n \exp(-x)$

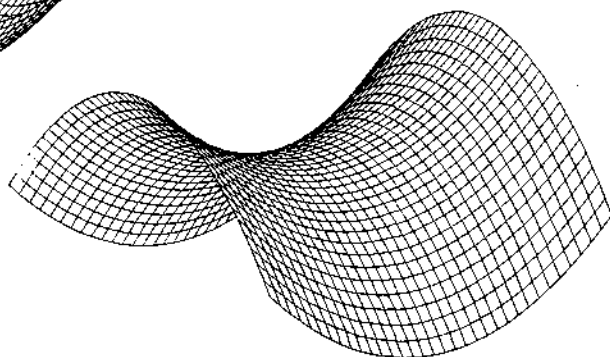
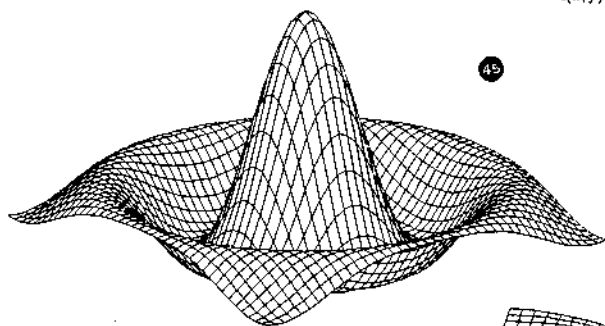
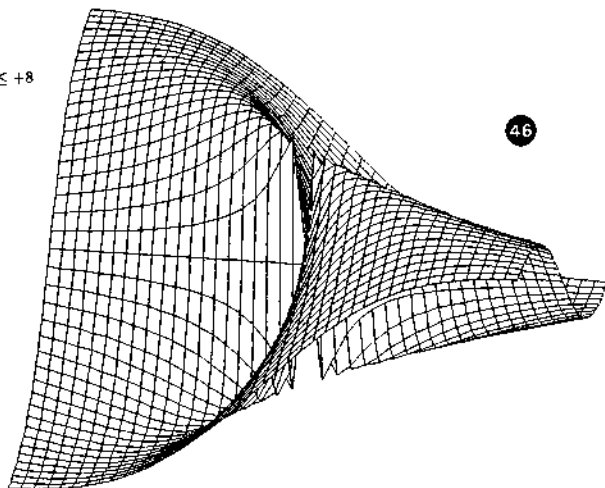
42

$f(x) = x^n \exp(-x^2)$

$f(x, y) = x^2 + y^2 \qquad -2 \le x, y \le +2$

$f(x, y) = x^2 - y^2 \qquad -5 \le x \le 5\ ,\ -6 \le y \le 6$

45

$f(x, y) = \dfrac{\sin\sqrt{x^2 + y^2}}{\sqrt{x^2 + y^2}} \qquad -8 \le x, y \le +8$

46

$f(x, y) = \dfrac{xy}{x^2 + y^2} \qquad -0.5 \le x, y \le 0.5$

INDICE ANALITICO